POLITECHNIKA WARSZAWSKA

DYSCYPLINA NAUKOWA AUTOMATYKA, ELEKTRONIKA, ELEKTROTECHNIKA I TECHNOLOGIE KOSMICZNE

DZIEDZINA NAUK INŻYNIERYJNO-TECHNICZNYCH

Rozprawa doktorska

mgr inż. Rafał Sikorski

Opracowanie metod monitorowania i kontroli procesu elektroerozyjnego drążenia mikrootworów

Promotor:

dr hab. inż. Krzysztof Siwek, prof. uczelni

Promotor pomocniczy:

dr inż. Bartosz Chaber

Warszawa 2024

Podziękowania

Dziękuję Zespołowi Rozwoju i Utrzymania Ruchu Laboratoriów oraz zespołowi Laboratorium Zaawansowanych Technologii Produkcyjnych za pomoc i fantastyczną współpracę podczas realizacji projektu.

Dziękuję Promotorowi za nieocenioną pomoc oraz motywację podczas pisania tej pracy.

STRESZCZENIE

Proces drążenia elektroerozyjnego (EDM – ang. Electrical Discharge Machining) to niekonwencjonalny proces obróbki, w którym usuwanie materiału odbywa się poprzez kontrolowane, powtarzalne wyładowania elektryczne. Ten proces jest znany ze swojej zdolności do obróbki „super stopów", które są trudne do przetwarzania tradycyjnymi metodami. Jedną z odmian procesu EDM jest proces elektroerozyjnego wiercenia otworów z użyciem elektrod w postaci cienkościennych rurek (FHD - ang. Fast Hole Drill). FHD jest obecnie technologią szeroko stosowaną w produkcji otworów wentylacyjnych chłodzenia filmowego części turbin silników odrzutowych. Potrzeba opracowania nowych technik usprawniających proces rozwojowy technologii drążenia otworów chłodzących oraz konieczność stworzenia nowych niezawodnych narzędzi diagnostyki i kontroli procesu produkcyjnego stanowi główny cel prac opisywanych w ramach niniejszej rozprawy.

FHD spełnia wysokie wymagania jakie stwarza stosowanie materiałów lotniczych, jednakże, ze względu na zmienny charakter procesu oraz skomplikowaną geometrię obrabianych części konieczne jest monitorowanie etapów oraz stabilności procesu, aby zapewnić bezawaryjne i pewne działanie. Jednym z głównych zagadnień w procesie jest wykrywanie przebicia. Zakończenie formowania otworu jest krytycznym etapem drążenia. Zbyt wczesne zatrzymanie wiercenia tworzy otwór o za małej średnicy, a tym samym nie spełniający wymagań przepływowych w procesie chłodzenia. Zbyt późne wykrycie może spowodować uszkodzenie powierzchni wewnętrznego kanału części turbiny, a tym samym zniszczenie elementu. Proces wykrywania jest niezwykle trudny ze względu na niejednostajne, zależne od wielu czynników, erodowanie części obrabianej i elektrody, co nie pozwala na jednoznaczne powiązanie kontroli etapu procesu wiercenia z danymi z maszyny jak np. przemieszczenie wrzeciona roboczego. Ze względu na wysoką zmienność procesu pomocne mogą być narzędzia pozwalające, przy niewielkim nakładzie pracy, na dogłębną analizę charakteru powstających wyładowań. Obecnie stosowane metody nie spełniają wysokich wymagań przemysłu lotniczego, z tego względu konieczne jest opracowanie nowych podejść wykorzystujących między innymi narzędzia sztucznej inteligencji i uczenia maszynowego.

Głównym celem niniejszej rozprawy jest analiza procesu FHD oraz stworzenie metodologii monitorowania przebiegu oraz kontroli stanu procesu i wykrywania anomalii w oparciu o narzędzia sztucznej inteligencji. Rozwiązania te pozwalają na wykrywanie

kluczowych etapów obróbki w czasie rzeczywistym, podczas trwania obróbki oraz rozpoznają potencjalne awarie obróbki lub niestabilności w procesie. W ramach rozprawy opracowano system monitorowania obróbki obejmujący czujniki rejestrujące wysokoczęstotliwościowe przebiegi prądu i napięcia oraz szereg parametrów rejestrowanych przez sterownik maszyny FHD. Ocena stanu procesu opiera się na analizie charakteru przebiegów w oparciu o narzędzia głębokiego uczenia, które pozwalają na autonomiczne rozpoznawanie charakteru procesu i na ich podstawie rozpoznają etap i prawidłowość prowadzenia wiercenia. Wydobyte cechy z zarejestrowanych przebiegów mogą dostarczyć istotnych informacji na temat zakłóceń w procesie lub potencjalnych uszkodzeń powierzchni obrabianych otworów.

W części eksperymentalnej pracy, z pomocą opracowanego systemu pomiarowego wykonano szereg drążeń testowych, które posłużyły jako dane wejściowe do pracowania inteligentnych narzędzi analizy i kontroli procesu FHD. Najpierw przeprowadzona została analiza offline, skupiającą się na badaniu procesu drążenia elektroerozyjnego w skali makro i mikro, w celu identyfikacji parametrów możliwych do zaobserwowania i mających duży wpływ na stabilność obróbki i identyfikacji przebicia ścianki detalu. Następnie opracowane zostało podejście pozwalające na automatyczne analizowanie przebiegu procesu ze względu na charakter przebiegu każdego z występujących impulsów EDM. Rozwiązanie posiada zdolność do autonomicznego klasyfikowania grup impulsów o różnych przebiegach w oparciu o nieprzetwarzane dane pomiarowe. Przetestowano różne podejścia oparte o metodologie ML (ang. Machine Learning) i DL (ang. Deep Learning) oraz porównano ich wyniki. Ostateczne narzędzie oparte o głębokie konwolucyjne sieci neuronowe pozwoliło na osiągnięcie dokładności klasyfikacji na poziomie przekraczającym 99,6% na danych testowych. Analiza przebiegów procesu wielu drążeń z wykorzystaniem zabudowanego klasyfikatora, pozwoliła na dobrą analizę procesu oraz identyfikowanie kluczowych zmian charakteru generacji impulsów w okolicy przełomu i przebicia.

Na podstawie wyciągniętych z dogłębnej analizy generacji impulsów EDM, dokonano próby zbudowania narzędzia wykrywania przebicia zdolnego do autonomicznego wykrywania przebicia na podstawie nieprzetworzonych sygnałów pomiarowych. W rozprawie przetestowano różne podejścia. Efektem prac było zbudowanie dwóch niezależnych modeli opartych na różnych podejściach: analizie całych pakietów surowych danych w oparciu o głębokie sieci CNN (ang. Convolutional Neural Network) i analizie częstości występowania poszczególnych grup impulsów w oparciu o sieć CNN-GRU (ang. Gated Recurrent Unit). Obie

testowane metody osiągnęły 100% trafność oraz dobrą terminowość w wykrywaniu momentu przebicia. Opracowane metody znacznie przewyższyły skuteczność wbudowanego algorytmu wykrywania przebicia elektrodrążarki i mogą stanowić lepszą alternatywę w warunkach produkcyjnych.

W ramach rozprawy opisana jest także metoda pozwalająca na pewne wykrywanie anomalii procesu, jedynie na podstawie surowych, wysokoczęstotliwościowych sygnałów napięcia i prądu generatora EDM. W pracy dokonana jest analiza różnych podejść oraz zaproponowany jest podejście rekonstrukcyjne oparte o głęboką sieć jednowymiarową CNN 1D. Zbudowany model pozwolił na wykrywanie symulowanych w części badawczej drążeń błędnych z dokładnością sięgającą 99,48%. Rozwiązanie jednoznacznie rozpoznaje dane wygenerowane w drążeniach testowych symulujących wystąpienie usterek na podstawie danych surowych z 20ms procesu, a także ze 100% pewnością potrafi rozpoznać drążenie poprawne wykonywane z różnymi zestawami parametrów.

Słowa kluczowe: wiercenie elektroerozyjne (FHD), monitorowanie procesu, sztuczne sieci neuronowe, uczenie maszynowe, analiza danych

ABSTRACT

The Electrical Discharge Machining (EDM) process is an unconventional machining process where material removal is achieved through controlled, repetitive electrical discharges. This process is known for its ability to machine "super alloys," which are difficult to process using traditional methods. One variant of the EDM process is the Fast Hole Drill (FHD), which involves drilling holes using thin-walled tube electrodes. FHD is currently a widely used technology for producing cooling film ventilation holes in jet engine turbine parts.

The need to develop new techniques to streamline the development process of cooling hole drilling technology and the necessity to create reliable diagnostic and control tools for the production process is the main goal of the work described in this dissertation.

FHD meets the high demands posed by the use of aerospace materials. However, due to the variable nature of the process and the complex geometry of the machined parts, it is necessary to monitor the stages and stability of the process to ensure reliable and safe operation. One of the main issues in the process is detecting breakthrough. The end of the hole-forming process is a critical stage of drilling. Stopping the drilling too early creates a hole with a diameter that is too small, failing to meet the flow requirements in the cooling process. Detecting too late can damage the surface of the inner channel of the turbine part, leading to part destruction. The detection process is extremely challenging due to the uneven erosion of the workpiece and electrode, which depends on many factors and does not allow for a clear connection between the control of the drilling stage and machine data such as the movement of the working spindle. Due to the high variability of the process, tools that allow for in-depth analysis of the nature of the discharges with minimal effort can be helpful. Currently used methods do not meet the high demands of the aerospace industry, making it necessary to develop new approaches utilizing artificial intelligence and machine learning tools.

The main objective of this dissertation is to analyze the FHD process and create a methodology for monitoring the progress and controlling the state of the process, as well as detecting anomalies based on artificial intelligence tools. These solutions enable the detection of key processing stages in real-time during machining and identify potential machining failures or process instabilities. As part of the dissertation, a machining monitoring system was developed, including sensors that record high-frequency current and voltage waveforms and a

range of parameters recorded by the FHD machine controller. The assessment of the process state is based on the analysis of the waveform characteristics using deep learning tools, which allow for autonomous recognition of the process nature and, based on this, identify the stage and correctness of the drilling. The extracted features from the recorded waveforms can provide significant information about process disturbances or potential surface damage of the machined holes.

In the experimental part of the work, a series of test drillings were carried out using the developed measurement system, which served as input data for developing intelligent tools for analyzing and controlling the FHD process. First, an offline analysis was conducted, focusing on studying the EDM process on a macro and micro scale to identify observable parameters that significantly affect machining stability and wall breakthrough detection. Then, an approach was developed that allows for automatic process analysis based on the nature of each EDM pulse. The solution could autonomously classify groups of pulses with different waveforms based on raw measurement data. Various approaches based on ML (Machine Learning) and DL (Deep Learning) methodologies were tested and their results compared. The final tool, based on deep convolutional neural networks, achieved classification accuracy exceeding 99.6% on test data. The analysis of the process waveforms from multiple drillings using the integrated classifier allowed for a good process analysis and identification of key changes in the nature of pulse generation around the breakthrough.

Based on the in-depth analysis of EDM pulse generation, an attempt was made to create a breakthrough detection tool capable of autonomously detecting breakthroughs based on raw measurement signals. Various approaches were tested in the dissertation. The result of the work was the development of two independent models based on different approaches: analyzing whole packets of raw data using deep CNN (Convolutional Neural Network) and analyzing the frequency of occurrence of individual pulse groups using the CNN-GRU (Gated Recurrent Unit) network. Both tested methods achieved 100% accuracy and good timeliness in detecting the breakthrough moment. The developed methods significantly outperformed the EDM machine's detection algorithm and can provide a better alternative in production conditions.

The dissertation also describes a methodology for reliably detecting process anomalies based solely on raw, high-frequency voltage, and current signals from the EDM generator. Various approaches are analyzed, and a reconstruction-based approach using a deep one-

dimensional CNN 1D network is proposed. The developed model allowed for the detection of simulated incorrect drillings in the experimental part with an accuracy of up to 99.48%. The solution unequivocally recognizes data generated in test drillings simulating defects based on raw data from 20ms of the process, and it can also distinguish correct drilling performed with different parameter sets with 100% certainty.

Keywords: Electrical Discharge Drilling (FHD), process monitoring, artificial neural networks, machine learning, data analysis

Spis treści

Opis skrótów i oznaczeń

AE – Autoenkoder (ang. Autoencoder)

AI – Sztuczna Inteligencja (ang. Artificial Intelligence)

ANN – Sztuczna sieć neuronowa (ang. Artificial neural network)

BCE – Binarna entropia krzyżowa (ang. Binary corssentropy)

CNC – Komputerowe sterowanie numeryczne (ang. Computerized Numerical Control)

CNN - Konwolucyjna sieć neuronowa (ang. Convolutional Neural Network)

CPU – Processor (ang. central processing unit)

DBN - Sieć głębokich przekonań (ang. Deep Belief Network)

DL – Uczenie głębokie (ang. Deep Learning)

DNN – Głęboka sieć neuronowa (ang. Deep Neural Network)

EDD – Drążenie elektroerozyjne (ang. Electroerosion Discharge Drilling)

EDD – Elektroerozyjne drążenie otworów (ang. Electrical Discharge Drilling)

EDM – Obróbka elektroerozyjna (ang. Electrical Discharge Machining)

FHD – Szybkie drążenie otworów EDM (ang. Fast Hole Drill)

GAF - Pole Kątowe Gramma (ang. Gramian Angular Field)

GAN – Generatywna sieć przeciwstawna (ang. Generative Adversarial Network)

GPU – Procesor graficzny (ang. graphics processing unit)

GRU – Bramkowana jednostka cykliczna (ang. Gated Recurrent Unit)

HMI – Aplikacja operatorska (ang. Human Machine Interface)

IoT – Internet Rzeczy (ang. Internet of Things)

kNN – K najbliższych sąsiadów (ang. K nearest neighbours)

ML – Uczenie maszynowe (ang. Machine Learning)

MSE - Błąd średniokwadratowy (ang. Mean Square Error)

MTF - Pole Przejść Markova (ang. Markov Transition Field).

OAA – Strategia jeden przeciw wszystkim (ang. one-against-all)

OAO – Strategia jeden przeciw jednemu (ang. one-against-one)

OC-SVM – Jednoklasowa metoda wektorów nośnych (ang. One Class Support Vector Machine)

PCA - Analiza składowych głównych (ang. Principal Component Analysis)

PGM - Probabilistyczny model graficzny (ang. Probabilistic Graphical Models)

RBF – Jądro funkcji radialnej (ang. Radial basis function kernel)

ResNet – Resztkowa sieć neuronowa (ang. Residual Neural Network)

RNN - Rekurencyjna sieć neuronowa (ang. Recurrent Neural Network)

SVDD – Wektory wspierające opisu danych (ang. Support Vector Data Description)

SVM – Metoda wektorów nośnych (ang. Support Vector Machine)

TET – Temperatura wlotowa turbiny (ang. Turbine Entry Temperature)

VAE – Autoenkoder wariacyjny (ang. Variational autoencoder)

WEDM – Cięcie elektroerozyjne drutowe (ang. Wire Electrical Discharge Machining)

WPT - Transformacja pakietów falkowych (ang. Wavelet Packet Transform)

WT – Transformacja fakowa (ang. Wavelet Transform)

1. Cel i teza rozprawy

Proces drążenia elektroerozyjnego otworów FHD (ang. Fast Hole Drill) ma ogromny potencjał w porównaniu z innymi tradycyjnymi technikami obróbki do obrabiania twardych komponentów metalowych. Szczególnie w produkcji silników lotniczych, gdzie konieczne jest zastosowanie wysokotemperaturowych i jednocześnie lekkich materiałów jak tzw. „super stopy" do produkcji łopatek kompresorów silnika odrzutowego, konieczne jest zastosowanie technologii spełniającej wysokie wymagania jakościowe, jaką jest proces FHD. Niemniej jednak, powtarzalność i efektywność procesu są ograniczone przez nieoczekiwane awarie procesu. W przeszłości wielu badaczy próbowało zrozumieć mechanizm przebiegu procesu drążenia EDM (ang. Electrical Discharge Machining) z wyszczególnieniem wykrywania przebicia oraz kontroli prawidłowości procesu, natomiast jeszcze nie zaproponowano rozwiązania dającego dostatecznie szybkie i niezawodne wyniki wymagane w procesie elektrodrążenia otworów wentylacyjnych. Dotychczasowe badania nie korzystały także w pełni z nowoczesnych metod przetwarzania danych, modelowania i wnioskowania jakie dają nam technologie oparte o uczenie maszynowe ML (ang. Machine Learning) i uczenie głębokie DL (ang. Deep Learning). Jak pokaże przeprowadzony przegląd literatury, wielu badaczy z sukcesem stosowało takie podejścia do analiz procesów technologicznych z wielu gałęzi techniki, osiągając często lepsze wyniki niż klasycznymi metodami kontroli. Fakt ten pozwala postawić tezę, że zastosowanie metod sztucznej inteligencji w procesie diagnostyki i kontroli przebiegu procesu FHD może znacznie usprawnić procesy badawczo rozwojowe tej technologii dając bardziej niezawodne wyniki przy jednoczesnym zmniejszeniu nakładu pracy ekspertów.

Przeprowadzony w ramach niniejsze rozprawy przegląd dotychczasowych prac badawczych opisujących proces FHD sugeruje, że system monitorowania stanu oparty na analizie wysokoczęstotliwościowych sygnałów elektrycznych z procesu i klasyfikacji impulsów EDM może być jednym z najbardziej niezawodnych sposobów przewidywania przebicia. Większość z proponowanych rozwiązań opiera się jednak na prostej identyfikacji impulsów w oparciu o wartości progowe, nie biorąc pod uwagę faktycznych kształtów poszczególnych impulsów. Podejście takie wymaga zaangażowania eksperckiego i nie jest odporne na zmienne parametry co zmniejsza jego niezawodność. Zbudowanie rozwiązania pozwalającego w pewny sposób automatycznie rozróżniać generowane wyładowania EDM, może posłużyć nie tylko jako silne narzędzie kontroli procesu, ale także może dostarczać

cennych, dogłębnych informacji o przebiegu procesu np. na potrzeby doboru optymalnych parametrów obróbki. Prace badawcze nie podejmowały także do tej pory prób zbudowania rozwiązania automatycznie rozpoznającego anomalie w przebiegach elektrycznych FHD. Narzędzie takie może być cenne ze względu na bieżącą kontrolę jakości wykonanych detali, a także może służyć do korelacji występujących defektów produkcyjnych z czynnikami je powodującymi. Zebrane dane o anomaliach drążeń, mogą być też dobrym źródłem danych wejściowych do tworzenia systemów predykcyjnego utrzymania ruch.

Tezą niniejszej rozprawy jest możliwość opracowania innowacyjnego podejścia wykorzystującego analizę sygnałów elektrycznych EDM oraz metody ML i DL do wykrywania przebicia i anomalii w procesie drążenia otworów wentylacyjnych metodą elektroerozyjną FHD.

Opracowanie metod integrujących zaawansowaną analizę sygnałów pochodzących z procesu drążenia z algorytmami uczenia głębokiego, umożliwi wczesne i niezawodne wykrywanie kluczowego etapu procesu jakim jest zakończenie formowania otworu i sygnalizowanie nieprawidłowości. Dzięki temu możliwe będzie zwiększenie precyzji i efektywności procesu FHD, a także minimalizacja ryzyka uszkodzeń i poprawa jakości końcowych produktów.

1.1. Cele Badawcze

Celem pracy badawczej było analiza przebiegu procesu oraz parametrów mierzonych mających kluczowy wpływ na prawidłowe identyfikowanie etapów i awarii procesu oraz zaproponowanie systemu monitorowania stanu i kontroli procesu dla elektroerozyjnego drążenia otworów wentylacyjnych w procesie FHD. W pracy opracowane zostały automatyczne narzędzia klasyfikujące poszczególne impulsy EDM. Stworzona i przetestowana była także metoda wykrywania właściwego momentu przebicia otworu oraz zaproponowane zostało podejście do analizy anomalii procesu na podstawie sygnałów elektrycznych z czujników. Opracowany system został zintegrowany z maszyną w środowisku produkcyjnym. Poniżej przedstawione są szczegółowe cele niniejszej pracy:

- Opracowanie metodologii akwizycji oraz przetwarzania danych pomiarowych procesu drążenia FHD,

- Analiza procesu drążenia otworów wentylacyjnych w procesie FHD. Identyfikacja możliwych do monitorowania parametrów najlepiej oddających prawidłowość przebiegu oraz etap zadania.
- Poznanie właściwości sygnałów prądu i napięcia zachodzących w czasie drążenia oraz określenie istotnych zmian przebiegów w zależących od etapu procesu.
- Przetestowanie różnych metod ML i DL owocujące zbudowaniem optymalnych modeli pozwalających na automatyczną klasyfikację impulsów.
- Analiza statystyczna wyodrębnionych i sklasyfikowanych impulsów oraz ewaluacja zmienności ich generacji w zależności od etapu procesu.
- Zbadanie wydajności różnych podejść i wyłonienie najlepszego rozwiązania do kontroli fazy procesu (w tym wykrycia przebicia) w oparciu o odczyty z czujników z użyciem podejścia DL.
- Eksperymentalne testy potwierdzające skuteczność zbudowanego podejścia kontroli procesu.
- Opracowanie algorytmu wykrywania błędów i anomalii procesu FHD w oparciu o sygnały elektryczne i metody ML i DL.

2. Wstęp

W dzisiejszych czasach wraz z dynamicznym rozwojem przemysłu coraz większe znaczenie zyskuje optymalizacja procesów produkcyjnych. Przemysł oraz środowiska naukowe prowadzą badania, które docelowo mają zwiększyć efektywność, redukować koszty oraz minimalizować zużycie zasobów. W kontekście tej ewolucji, technologie wytwarzania muszą dostosowywać się do wymogów nowoczesnego środowiska produkcyjnego, zwanego przemysłem 4.0, który charakteryzuje się intensywnym wykorzystaniem automatyzacji, cyfryzacji oraz analizy danych.

Przemysł lotniczy od zawsze był jedną z gałęzi wiodących prym na polu rozwoju i implementacji nowych technologii produkcyjnych w tym także procesów transformacji cyfrowej. Jest to sektor charakteryzujący się wysokim stopniem zaawansowania technologicznego, innowacyjnością i surowymi wymogami dotyczącymi jakości oraz bezpieczeństwa. Stanowi kluczowy filar transportu pasażerskiego i towarowego na skalę globalną, a także odgrywa istotną rolę w celach militarnych. Jedną z nieodzownych części przemysłu lotniczego jest rozwój i produkcja silników, w szczególności szeroko stosowanych w dzisiejszych czasach różnych odmian silników odrzutowych jak silniki turboodrzutowe, turbośmigłowe, turboprop i bezśmigłowe. Silniki odrzutowe od swoich początków za czasów 2 wojny światowej rozwijają się w szybkim tempie. Wciąż trwają intensywne prace mające na celu zwiększanie efektywności paliwowej, redukcję emisji spalin oraz hałasu, zwiększanie osiągów i mocy, zastosowanie alternatywnych paliw czy integrację z systemami elektrycznymi. Aby umożliwić rozwój konieczne, jest stałe wprowadzanie nowych technologii i materiałów, co często wymusza zastosowanie niestandardowych metod produkcyjnych.

Jednym z kluczowych procesów w produkcji silników odrzutowych, który wymaga precyzji i zaawansowanych technologii, jest drążenie otworów wentylacyjnych chłodzenia filmowego. Chłodzenie wysokotemperaturowych elementów turbin jest niezbędną częścią technologii nowoczesnych silników, ze względu na ich pracę w warunkach często przekraczających wytrzymałość temperaturową samych materiałów. W tym kontekście, metoda elektroerozyjna (EDM) wyłoniła się jako niezwykle skuteczne narzędzie, umożliwiające precyzyjne wykonywanie otworów o skomplikowanych kształtach w różnych materiałach, w tym w stopach metali (superstopach) używanych w produkcji wysokotemperaturowych komponentów silników odrzutowych.

W niniejszej rozprawie analizie poddany został proces drążenia elektroerozyjnego otworów wentylacyjnych w kontekście przemysłu 4.0 [1-3] oraz ideologii Lean. W pracy opisane zostały opracowane narzędzia monitorowania, kontroli i analizy przebiegu procesu oparte o mechanizmy uczenia maszynowego oraz uczenia głębokiego wdrożone w celu optymalizacji wydajności i minimalizacji odpadów. Jednym z głównych opisywanych zagadnień jest nowatorskie podejście do automatycznego wykrywania przebicia w procesie drążenia otworów, które stanowi kluczowy element zapewnienia jakości i bezpieczeństwa wytwarzanych komponentów. Błędne wykrywanie przebicia jest obecnie jednym z głównych powodów prowadzącym często do zniszczenia części i generowania przez to dużych kosztów. Dlatego też skuteczne metody wykrywania tego zjawiska są niezwykle istotne dla procesu produkcyjnego. Ponadto opracowane metody pozwolą na klasyfikację impulsów EDM zachodzących podczas wiercenia oraz wykrywanie anomalii procesu na podstawie zarejestrowanych wysokoczęstotliwościowych przebiegów prądu i napięcia. Podejście takie pozwoli na znacznie dokładniejszą analizę warunków i parametrów przebiegu procesu. Opracowana metodyka zapewnia możliwość precyzyjnego określania warunków kształtowania otworów, co otwiera nowe możliwości np. do uzyskania korelacji wydajności drążenia i jakości wykonywanych otworów z parametrami procesu. Dodatkowo, szczegółowa analiza sygnałów prądu i napięcia pozwala na łatwiejsze identyfikowanie warunków i przyczyn powstawania defektów produkcyjnych oraz może zostać zastosowana w predykcyjnej kontroli stanu maszyny

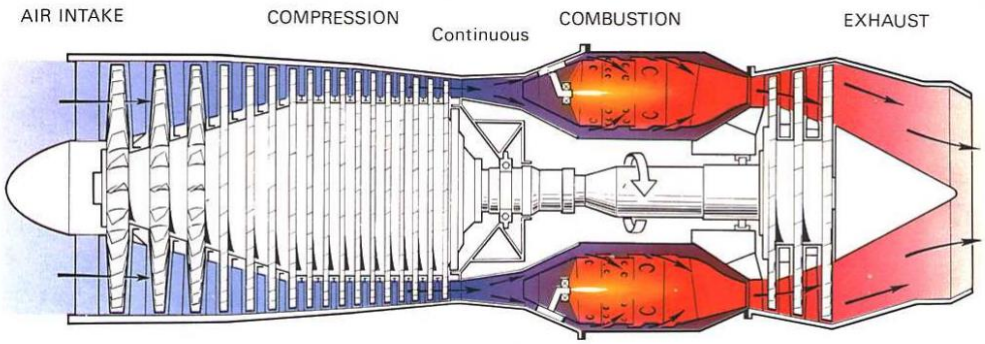
Poprzez połączenie zaawansowanych technologii drążenia elektroerozyjnego z koncepcjami przemysłu 4.0 [1-13], ideologią Lean [14-18] oraz metodyką tworzenia opartych na danych inteligentnych systemów produkcyjnych [19-25], rozprawa prezentuje kompleksowe podejście do produkcji otworów wentylacyjnych w silnikach odrzutowych. Podejście to pozwoli na osiągnięcie najwyższej jakości przy minimalnym zużyciu zasobów oraz maksymalnej efektywności procesu produkcyjnego.

Niniejsza praca skupiać się będzie na opracowaniu opartych o dane inteligentnych rozwiązań do monitorowania i kontroli procesu elektroerozyjnego wiercenia otworów do chłodzenia filmowego części turbin silników odrzutowych. Technologia ta jest nieodzownym elementem w produkcji współczesnych silników odrzutowych. W kolejnym rozdziale przybliżona będzie idea działania silników odrzutowych oraz opisane będą czynniki,

ze względu na które konieczne jest stosowanie niestandardowych procesów do obróbki, takich jak EDM w ich produkcji.

2.1. Silniki odrzutowe

Technologia silników odrzutowych jest rozwijana od początku lat 40-tych XX wieku [26] kiedy po raz pierwszy zastosowano silnik tego typu do napędzania samolotów, takich jak Gloster E28/39 i Junkers Jumo 004. Projektantami i pomysłodawcami technologii byli Sir Frank Whittle i Anselm Franz. Ich projekty dały podwaliny nowoczesnym silnikom odrzutowym, które w dużej mierze nadal opierają się na tych samych zasadniczych działania. Od lat 50-tych XX wieku rozpoczął się okres intensywnego rozwoju i konkurencji w branży silników odrzutowych, który trwa nieprzerwanie do dziś. Nowoczesne silniki składają się z tysięcy elementów i podlegają ciągłym analizom, optymalizacjom i zmianom. Ogólny, ideowy schemat działania silnika turbo-odrzutowego przedstawiony został na rysunku 1.



Rysunek 1 Schemat budowy osiowego silnika turbo-odrzutowego [26]

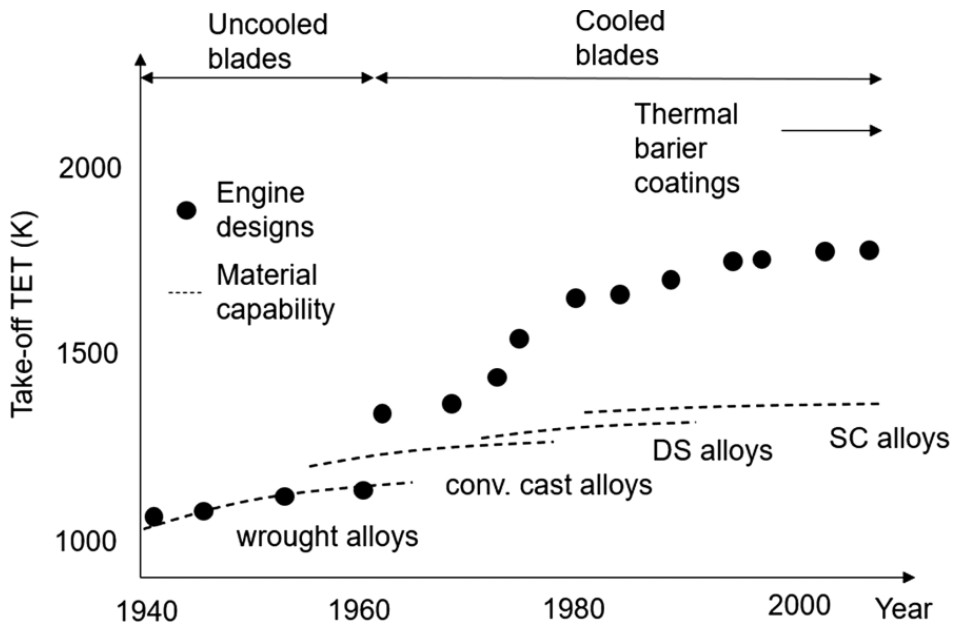
Silnik turbinowy to zasadniczo silnik cieplny, wykorzystujący powietrze jako czynnik roboczy do generowania ciągu. Aby to osiągnąć, powietrze przechodzące przez silnik musi być przyspieszone; oznacza to zwiększenie prędkości lub energii kinetycznej powietrza. W celu uzyskania tego wzrostu, najpierw zwiększa się energię ciśnienia, następnie dodaje się energię cieplną, zanim ostatecznie przekształci się ją z powrotem w energię kinetyczną w postaci strumienia gazów o dużej prędkości [26]. Łopatki turbiny odbierają energię generowaną w komorze spalania od gazów płynących kanałem obejmującym turbinę i wylot, tym samym napędzając ruch obrotowy wału silnika. Cześć energii jest przekazywana na potrzeby kompresora, reszta generuje ciąg silnika.

Konkurencja na rynku lotniczym wymusza stały pościg za osiąganiem coraz większej wydajności i niezawodności z zachowaniem bardzo wysokich wymagań dotyczących bezpieczeństwa użytkowania. Podstawowym kierunkiem rozwoju pozwalającym na poprawę wydajności silnika jest podnoszenie temperatury cyklu pracy. Elementy współczesnych silników muszą być wykonane z materiałów oraz muszą posiadać rozwiązania technologiczne pozwalające na długotrwałą pracę w temperaturach rzędu 1350-1500°C [27]. Aby umożliwić pracę w tak ekstremalnych warunkach konieczne jest zastosowanie odpowiednich materiałów, powłok antykorozyjnych, ceramicznych barier termicznych i technik chłodzenia.

W celu spełnienia tych wymagań do produkcji elementów narażonych na wysokie temperatury stosować zaczęto tzw. „super stopy" po raz pierwszy opracowane w latach 20 XX wieku [28]. Materiały te zapewniają odporność na trudne środowisko pracy, dużą wytrzymałość wysoko i niskocyklową oraz mają dużą żaroodporność, tzn. odporność na zerwanie w wyniku pełzania.

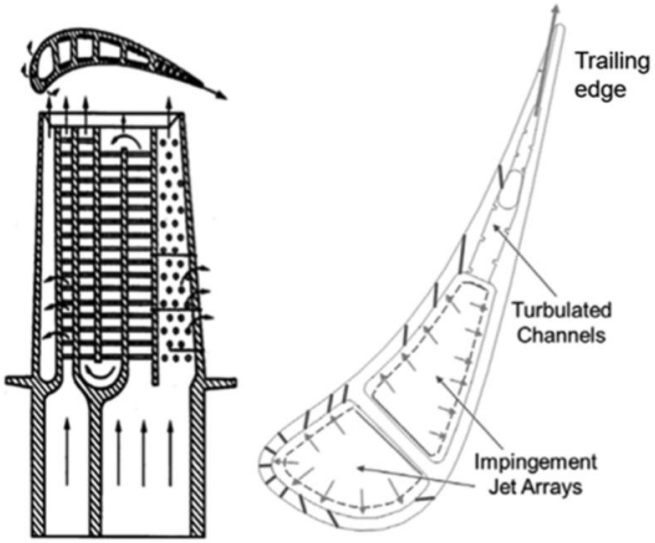
Zastosowanie tych materiałów pozwoliło na poprawę parametrów pracy silników, ale jednocześnie wymusiło wprowadzenie nowych technologii ich obróbki.

Rysunek 2 [29] obrazuje wzrost temperatury wlotowej turbiny TET (ang. Turbine Entry Temperature) w kolejnych generacjach silników oraz możliwości temperaturowe materiałów. Jak widać, w niektórych sekcjach turbiny temperatura może przekroczyć zdolność cieplną materiałów używanych do produkcji jej części. Z tego względu konieczne jest stosowanie dodatkowych rozwiązań technologicznych, które pozwolą na długotrwałą pracę silnika. Chłodzenie elementów turbiny, które są najbardziej narażone na obciążenia cieplne, to główna technika wykorzystywana do pokonania tej przepaści i stanowi zatem kluczowe wymaganie dla zwiększenia efektywności i trwałości turbiny gazowej.



Rysunek 2 Wzrost temperatury pracy silników Rolls-Royce [29]

Jedną z podstawowych metod zapewnienia odpowiednich warunków pracy silnika jest wytwarzanie otworów chłodzących elementy turbiny, które generują efekt chłodzenia filmowego na ich powierzchni. W swojej najbardziej podstawowej formie, ta metoda tworzy warstwę chłodziwa (film chłodzący) oddzielającą ścianę komponentu od głównego kanału gazowego. W praktycznych zastosowaniach, uwzględniając integralność strukturalną części, jednolite szczeliny są rzadko możliwe, a typowe rozwiązania obejmują dyskretyzację szczeliny na rzędy otworów chłodzących. Na rysunku 3 przedstawiono przykładowe przekroje chłodzonych elementów turbiny [30].



Rysunek 3 Przekrój chłodzonej łopatki i kierowniczy turbiny [30]

Mechanizm chłodzenia filmowego uważany jest za kluczową część nowoczesnego projektu turbiny. Pozwala na zmniejszenie strumienia ciepła do części, zamiast zwiększania ekstrakcji ciepła.

W poprzednich generacjach silników lotniczych otwory chłodzące wykonywane były w procesie tradycyjnego wiercenia mechanicznego lub poprzez obróbkę elektrochemiczną. Wprowadzenie „super stopów" o dużej twardości oraz ciągłe zmniejszanie średnic otworów wymusiło odejście od tych metod i stosowanie innych technologii. Rozwój technologiczny w dziedzinie układów tranzystorowych i wprowadzenie obrabiarek sterowanych numerycznie pozwoliły na zastosowanie obróbki elektroerozyjnej do drążenia otworów chłodzących. Proces wykonywania otworów polega na użyciu elektrody o kształcie cienkościennej rurki, która wykonana jest z materiału przewodzącego. Metoda obróbki elektroerozyjnej pozwala także na wykonywanie bardziej skomplikowanych kształtowych otworów chłodzących. Otwory tego typu pozwoliły na dalsze polepszenie parametrów chłodzenia i w efekcie podniesienie temperatury cyklu pacy i efektywności całego silnika [31].

Na przestrzeni lat przebieg podstawowego cyklu pracy silników odrzutowych nie uległ zasadniczej zmianie, natomiast same parametry pracy zmieniły się dramatycznie. Jak pokazuje rysunek 3, temperatura TET uległa podwojeniu, z wartości 1000°C do prawie 2000°C w nowoczesnych konstrukcjach. Ciągłe poprawianie parametrów pracy wymusza także wprowadzanie coraz to bardziej skomplikowanych procesów produkcyjnych. Ponieważ produkcja elementów silników odrzutowych jest procesem bardzo kosztownym, wielkie znaczenie, oprócz ulepszania samej technologii silnika, ma też wprowadzanie rozwiązań mających potencjał ograniczenia kosztów oraz wykrywania defektów produkcyjnych. Fakt ten jest główną motywacją towarzyszącą tworzeniu niniejszej rozprawy.

2.2. Obróbka elektroerozyjna – analiza zagadnienia

Część eksperymentalna pracy dotyczyć będzie monitorowania oraz tworzenie rozwiązań do diagnostyki i kontroli procesu elektroerozyjnego drążenia otworów wentylacyjnych elementów turbin silników odrzutowych. Rozwój silników lotniczych wprowadził w ich konstrukcji stosowanie tzw. superstopów ze względu na ich doskonałą odporność na pełzanie, korozję i utlenianie, a także bardzo wysoką wytrzymałość i twardość przy podwyższonych temperaturach [32, 33]. Stosowanie superstopów, pozwoliło

na osiągnięcie dużych korzyści w konstrukcji części turbin, co pociągnęło za sobą konieczność wprowadzenia nowych technologii i rozwiązań koniecznych w procesie ich produkcji. Superstopy są bowiem materiałami trudnymi do obróbki, charakteryzują się właściwościami niekorzystnymi dla procesu obróbki skrawaniem: słabą przewodnością cieplną, wysoką wytrzymałością przy podwyższonych temperaturach, wysokim współczynnikiem hartowania przy pracy i silnym powinowactwem chemicznym do materiałów narzędziowych [34]. Te właściwości powodują częste uszkodzenia narzędzi skrawających, takie jak ścieranie brzegu, ścieranie powierzchni bocznej, kruszenie itp. [35], co skutkuje szybkim zużywaniem się narzędzia obróbczego i zwiększają naprężenia i temperaturę na powierzchni narzędzia, przez co pogarszają jakość wykończonej powierzchni [36].

Wymagania związane z wprowadzeniem do użytku nowych materiałów pociągnęły za sobą wiele różnych kierunków rozwoju rozwiązań obróbki superstopów. Przykładami mogą być prace w kierunku rozwoju obróbki na gorąco z wykorzystaniem podgrzewania laserowego [37], obróbki kriogenicznej [38], obróbki obrotowej [39], obróbki hybrydowej [40] oraz obróbki z użyciem chłodziwa o wysokim ciśnieniu [41].

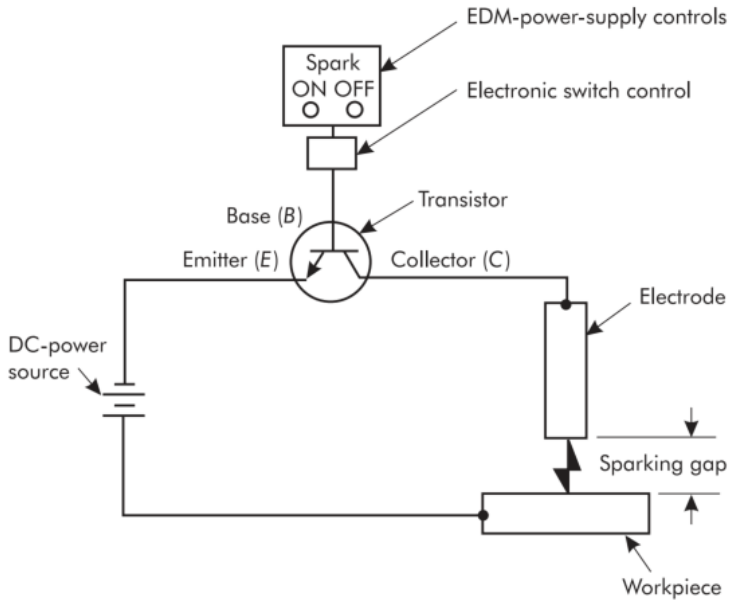
W technologii produkcji części turbin silników lotniczych takich jak łopatki i kierownice turbin, wiodącym rozwiązaniem jest stosowanie obróbki elektroerozyjnej. W tym rozdziale opisane zostały podstawy tej metody obróbki.

Historyczny rys rozwoju technologii elektroerozyjnej dobrze opisał w swojej pracy Schumacher [43]. Proces obróbki elektroerozyjnej ma swoje początki już w 1770, kiedy to Joseph Priestley zaobserwował jako pierwszy erozję metalu z pomocą ładunku elektrycznego. Efekt ten został po raz pierwszy celowo zastosowany do obróbki elektroerozyjnej i opisany przez Borysa i Natalię Łazarienko w 1943 roku. Ich odkrycie dało podwaliny do rozwoju technologii EDM.

Przełomem w zastosowaniu technologii EDM było wynalezienie układów tranzystorowych. W obecnie stosowanych maszynach, układy tranzystorowe odpowiedzialne są za napięcie, natężenie oraz częstotliwość generowania wyładowań. Rysunek 4 pokazuje uproszczoną ideę działania takiego układu.

Kolejnym dużym krokiem mającym wpływ na obecny kształt maszyn EDM było użycie roboczych osi posuwu sterowanych numerycznie. Pierwsze takie urządzenie zostało

po raz pierwszy skonstruowane w Związku Radzieckim oraz zaprezentowane na targach Expo w Montrealu w 1967 roku. Wszystkie obecnie konstruowane maszyny EDM opierają się na tych samych zasadniczych regułach działania.



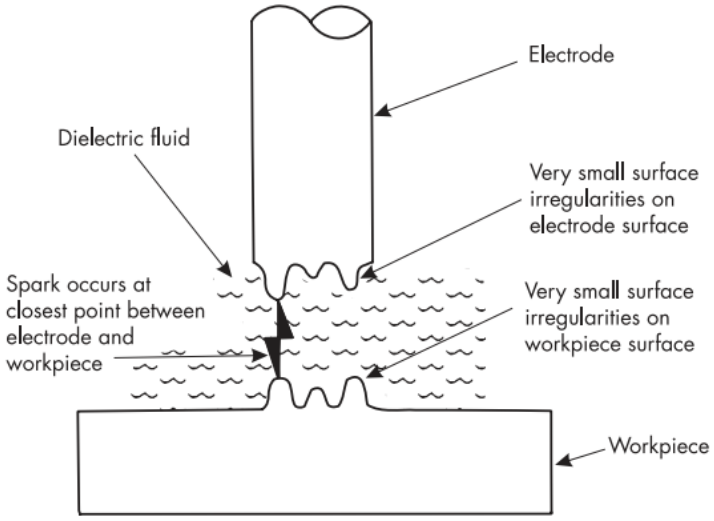
Rysunek 4 Uproszczony tranzystorowy układ generowania iskier [42]

Obróbka elektroerozyjna EDM (ang. Electrical Discharge Machining) to proces obróbki ubytkowej materiałów przewodzących prąd elektryczny, wykorzystujący precyzyjnie kontrolowane iskry, które występują pomiędzy elektrodą, a przedmiotem obrabianym w obecności płynu dielektrycznego. Elektrodę można porównać tutaj do narzędzia tnącego w obróbce skrawaniem.

Proces EDM często nazywany jest obróbką niestandardową, ponieważ narzędzie nie ma bezpośredniego kontaktu z elementem obrabianym. W czasie obróbki elektroda oddalona jest od elementu obrabianego o odległość pozwalającą na tworzenie iskier. Co za tym idzie w czasie obróbki nie występuje siła działająca na narzędzie.

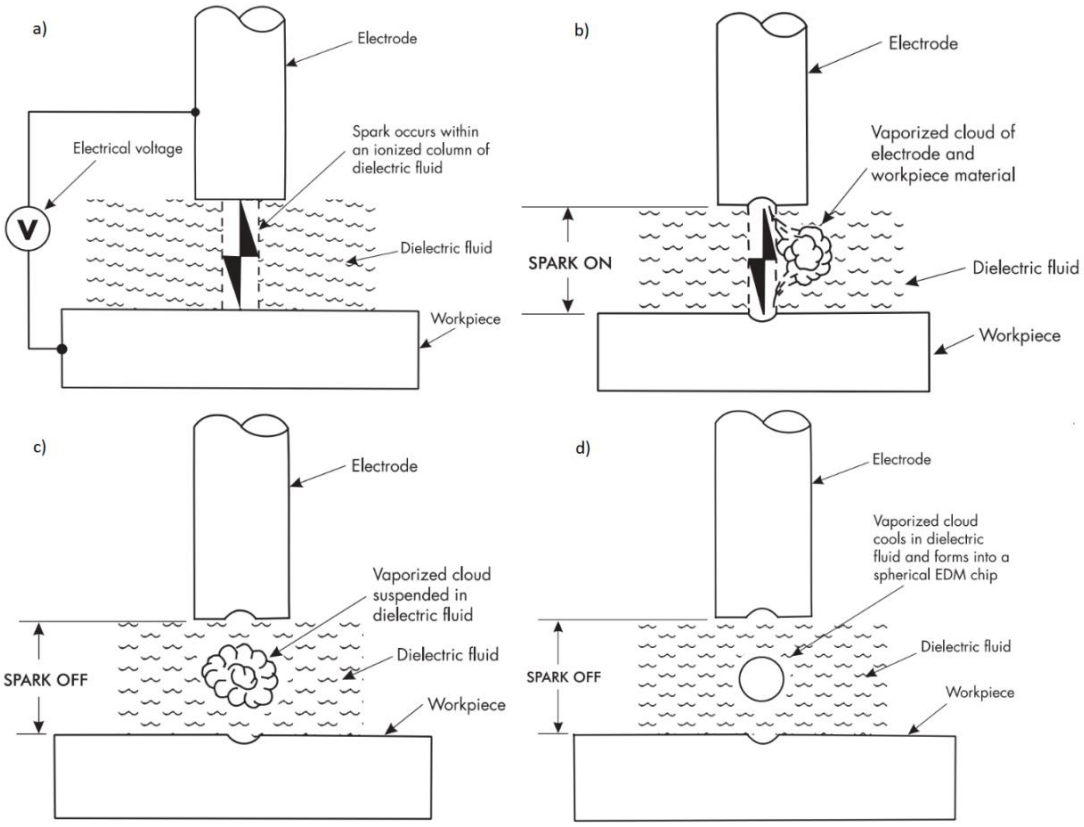
Ideę podstawowych zjawisk zachodzących podczas procesu EDM dobrze opisał Elman [42] w swojej pracy „Electrical Discharge Machining”. Elman opisuje, że podstawową regułą procesu EDM jest fakt, że tylko jedna iskra jest generowana w danym momencie. W czasie procesu, iskrzenie zachodzi z bardzo dużą częstotliwością od 2000 do 500 000 iskier na sekundę, co może sprawiać wrażenie występowania wielu wyładowań zachodzących

jednocześnie. Generowanie iskry zachodzi pomiędzy najbliższymi punktami elektrody i elementu obrabianego jak pokazano na poniższym rysunku 5.



Rysunek 5 Tworzenie iskry między elektrodą, a obiektem obrabianym [42]

W procesie EDM potencjał elektryczny zostaje przyłożony pomiędzy elektrodę, a element obrabiany. Pomiędzy elektrodą, a obiektem znajduje się izolator elektryczny, którego własnością jest możliwość zmiany charakteru na przewodzący, pod wpływem dostatecznie wysokiego napięcia elektrycznego. Płyn stosowany w procesie zachowuje swoje własności izolujące do czasu zbliżenie elektrody na dostatecznie małą odległość do obiektu (Rysunek 6 a)). Punkt zmiany charakteru płynu dielektrycznego nazywany jest punktem jonizacji. Formowanie kanału jonowego jest inicjowane przez pole elektromagnetyczne tworzące ścieżkę jonową umożliwiającą transport elektronów. Przez ścieżkę jonową zaczynają przepływać elektrony. W tym czasie natężenie prądu rośnie, a napięcie spada. W ten sposób powstaje iskra jako ścieżka elektronowa, łącząca elektrodę z materiałem detalu obrabianego. EDM jest procesem cieplnym. Materiał znajdujący się w pobliżu miejsca tworzenia iskry na elektrodzie i części obrabianej jest podgrzewany do temperatury pozwalającej na odparowanie materiału (Rysunek 6 b)). Odparowany materiał tworzy chmurę gazu uwięzioną wewnątrz płynu izolującego (Rysunek 6 c)).



Rysunek 6 Przebieg procesu EDM [42]

Kiedy impuls elektryczny zostaje wyłączony kanał plazmowy zapada się, natężenie spada do zera, a napięcie rośnie. Płyn odzyskuje własności dielektryczne, a para w nim uwięziona krzepnie tworząc tzw. wiór EDM, bardzo małą, pustą sferę stworzoną z materiału elektrody i obiektu obrabianego (Rysunek 6 d)). Dla poprawnego prowadzenia procesu powstały wiór powinien być skutecznie usuwany ze strefy obrabianej. Dokonuje się tego najczęściej poprzez stałe płukanie szczeliny iskrowej płynem dielektrycznym. Cykliczne powtarzanie procesu tworzenie iskier z bardzo dużą częstotliwością powoduje obserwowalną erozję detalu obrabianego oraz zużycie elektrody.

Obecnie stosowane rozwiązania obróbki elektroerozyjnej można podzielić na trzy podstawowe typy: wgłębne drążenie elektroerozyjne (ang. Sinker EDM), cięcie elektroerozyjne drutowe WEDM (ang. Wire Electrical Discharge Machining) oraz elektroerozyjne wiercenie otworów FHD (ang. Fast Hole Drilling).

W maszynach do cięcia drutowego WEDM jako elektrodę wykorzystuje się ciągły drut. Iskrzenie następuje od powierzchni drutu elektrodowego do przedmiotu obrabianego. Maszyny do drążenia wgłębnego wymagają elektrody o kształcie odwrotnym do kształtu uzyskiwanym

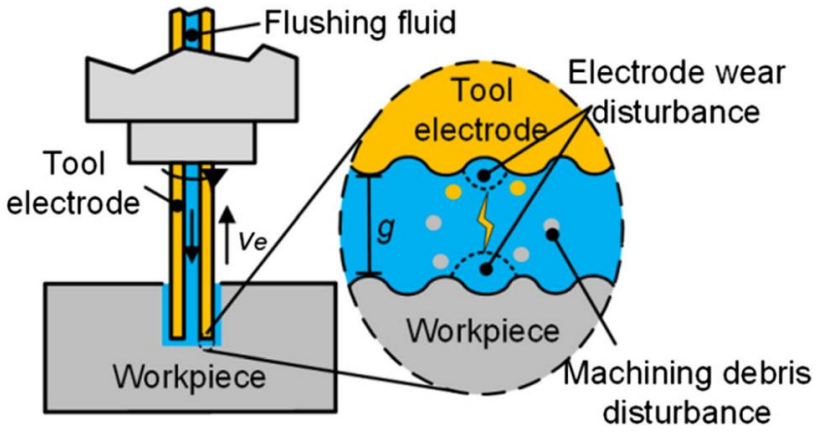
w przedmiocie obrabianym. Zagłębianie kształtowej elektrody powoduje odtworzenie jej kształtu w detalu obrabianym. Pewną ewolucją kształtowego drążenia elektroerozyjnego jest technologia elektroerozyjnego drążenia otworów EDD (ang. Electrical Discharge Drilling) lub inaczej FHD (ang. Fast Hole Drilling).

Część badawcza pracy będzie koncentrowała się na procesie elektroerozyjnego drążenia FHD. W następnym rozdziale opisane zostały zasady działania oraz najistotniejsze zagadnienia dotyczące tego procesu.

2.3. Elektrodrążenie małych otworów (wiercenie elektroerozyjne)

Drążenie elektroerozyjne otworów FHD (ang. Fast Hole Drill) lub inaczej EDD (ang. Electrical Discharge Drilling) po raz pierwszy pojawiło się w latach 80 XX wieku. Technologia ta pozwala na wykonywanie otworów przy użyciu procesu EDM. Wcześniej do wykonywania otworów w detalach obrabianych stosowane były standardowe maszyny wgłębne wyposażone w elektrody grzebieniowe. Technologia ta posiadała niestety wiele wad. Elektroda grzebieniowa jest bardzo wrażliwa na uszkodzenia i jej ogólny stan mocno negatywnie przekładał się na jakość wykonywanych otworów. Metoda ta wymaga stosowania płukania zewnętrznego co w wypadku małych średnic otworów generowało problemy. Przepływ dielektryka wprawiał elektrodę w drgania co pogarszało jakość otworów. Pomimo możliwości wiercenia wielu otworów jednocześnie, technologia grzebieniowa nie osiągała szybkich czasów produkcji. Nierównomierne wypalanie się poszczególnych zębów grzebienia elektrody wymuszało częste przerywanie procesu w celu jej regeneracji.

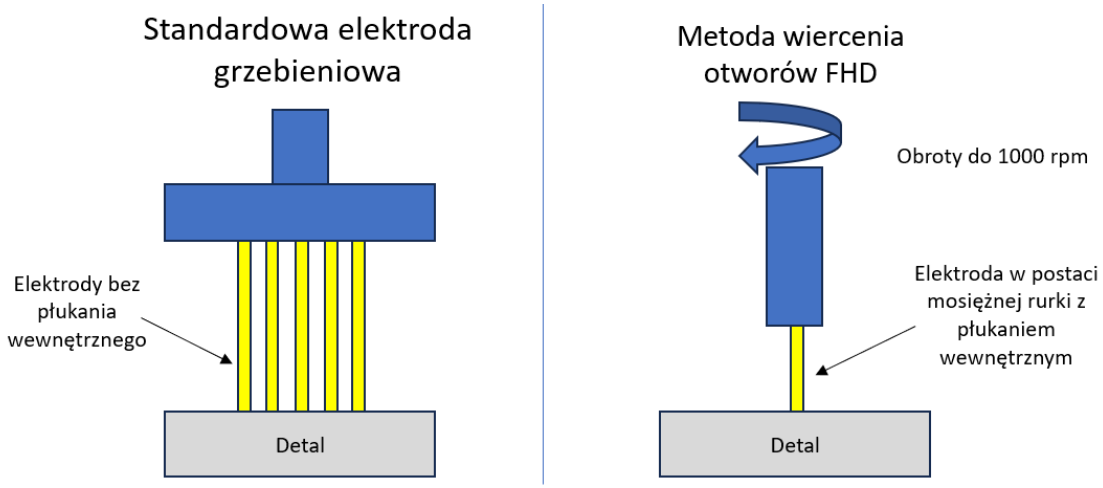
Jako alternatywę stosuje się obecnie drążenie pojedynczych otworów elektrodą rurkową w procesie Fast Hole Drill. Rysunek 7 pokazuje ideę procesu FHD.



Rysunek 7 Schemat procesu FHD [44]

W procesie tym elektroda w postaci cienkościennej rurki zagłębiana jest w obiekt obrabiany z wykorzystaniem tego samego zjawiska co we wgłębnym EDM. Aby stabilizować elektrodę, jest ona stale wprawiana w ruch obrotowy do prędkości ok. 1000 obr/min. Przez elektrodę pod ciśnieniem podawany jest dielektryk, który ma za zadanie wypłukiwać roztopione cząsteczki materiału i wióry EDM z przestrzeni roboczej. Jako płyn płukania w FHD stosuje się zwykle wodę demineralizowaną.

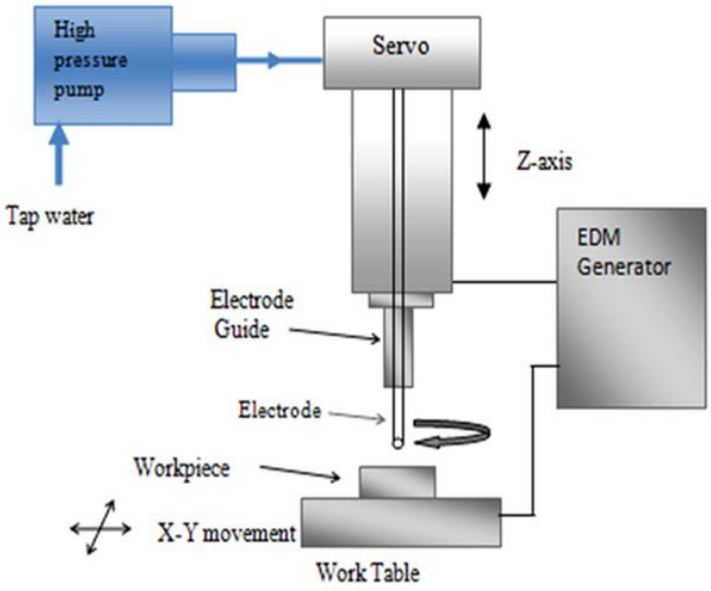
Proces FHD okazał się znacznie bardziej wydajny od drążenia grzebieniowego. Samo drążenie jest szybsze oraz pozwala na dokładniejsze pozycjonowanie otworów. W FHD położenie wykonywanego otworu kontrolowane jest przez trajektorię głowicy roboczej maszyny CNC zamiast przez kształt samej elektrody. Podstawowe różnice w podejściach do drążenia otworów procesem EDM zobrazowano na rysunku 8.



Rysunek 8 Porównanie drążenie otworów metodą grzebieniową oraz elektrodą rurkową

Kolejną zaletą FHD jest sam koszt procesu. Elektroda rurkowa jest znacznie tańsza w porównaniu do skomplikowanej w swojej konstrukcji elektrody grzebieniowej. Stosowanie ruchu obrotowego eliminuje drgania wywoływane przepływem dielektryka. Podawanie płynu przez samą elektrodę bezpośrednio do miejsca występowania iskier bardzo usprawnia usuwanie zanieczyszczeń, których nagromadzenie może być przyczyna usterek procesu.

Obecne maszyny FHD budowane są w postaci wieloosiowych obrabiarek sterowanych numerycznie, często posiadających nawet 5 niezależnych osi sterowanych. Podstawową budowę takiej maszyny przedstawiono na poniższym rysunku 9.

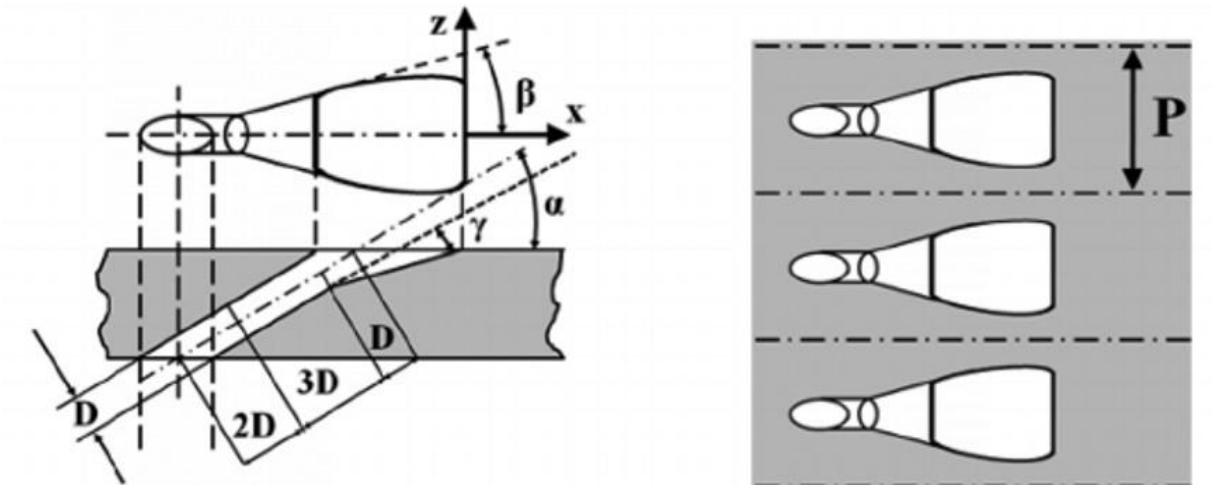


Rysunek 9 Podstawowa budowa maszyny FHD [45]

Cześć obrabiana zamontowana jest na stole roboczym i jest w całości zanurzona w zbiorniku z dielektrykiem, zwykle wodą destylowaną. Stół roboczy posiada możliwość operowania detalem obrabianym w polu roboczym, co pozwala na wykonywanie wielu otworów w różnych częściach detalu. Nad stołem roboczym znajduje się układ wrzeciona elektrody. Elektroda prowadzona jest przez przystosowaną do jej rozmiarów prowadnicę, która stabilizuje jej ruch. Przez cały czas obróbki elektroda wprawiana jest w ruch obrotowy wokół własnej osi. Całość wrzeciona, na którym zamontowana jest elektroda przemieszczana jest w osi Z za pomocą dedykowanego serwomechanizmu EDM. Przesuwanie elektrody w osi Z pozwalana na stopniowe zagłębianie jej w detalu obrabianym i w efekcie drążenie otworu procesem EDM. Pomiędzy elektrodą, a stołem roboczym włączony jest generator EDM, który jest odpowiedzialny za wysokoczęstotliwościowe tworzenie iskier w szczelinie między

elektrodą, a przedmiotem obrabianym. Środkiem elektrody stale tłoczony jest dielektryk do szczeliny EDM. Za ten układ odpowiedzialna jest dedykowana pompa wysokiego ciśnienia. W standardowym procesie obróbki tworzone są całe serie otworów chłodzących. Po zakończeniu tworzenia danego otworu wrzeciono wycofywane jest wzdłuż osi Z, stół przemieszcza część w położenie kolejnego otworu. Proces jest kontynuowany w trybie automatycznym do utworzenia wszystkich otworów w obrabianym detalu.

Elektrodę rurkową, oprócz drążenia prostych otworów okrągłych można stosować do wykonywania otworów kształtowych. Elektrodę można stosować podobnie do freza palcowego przesuwając elektrodę nad materiałem obrabianym po określonej trajektorii. Możliwość ta jest niezwykle przydatna podczas produkcji otworów do chłodzenia filmowego części turbin.



Rysunek 10 Przykładowy kształtowy otwór chłodzący (26)

Liczne publikacje [46, 47] potwierdzają wyższość otworów z zakończeniem kształtowym (Rysunek 10) zarówno w zakresie osiąganych efektów chłodzenia jak i pod względem strat aerodynamicznych.

Proces tworzenia filmu chłodzącego na elementach turbiny wymaga podania powietrza do środka elementu i wypuszczenie go na zewnątrz wykonanym otworem. Ze względów wytrzymałościowych, warunków chłodzenia oraz wymagań konstrukcyjnych silników jak np. konieczność ograniczania masy, wewnętrze powierzchnie elementów często posiadają skomplikowaną geometrię. Podczas wiercenia otworów wewnętrzne powierzchnie nie są widoczne. Prawidłowe wykonanie otworu wymaga zatrzymania procesu zaraz po całkowitym przebiciu wierconej ścianki. Niedopuszczalne jest rozpoczęcie drążenia

w ściance przeciwległej. Trzeba pamiętać, że równie groźne jest zbyt wczesne zatrzymanie procesu. Wykonie nie w pełni przelotowego otworu ogranicza przepływ gazu chłodzącego przez co nie może zostać utworzony film chłodzący i część silnika może ulec szybkiemu zniszczeniu termicznemu. Te ograniczenia wymuszają wprowadzenie metod i narzędzi pozwalających na monitorowanie procesu i eliminowanie tego typu błędów obróbki.

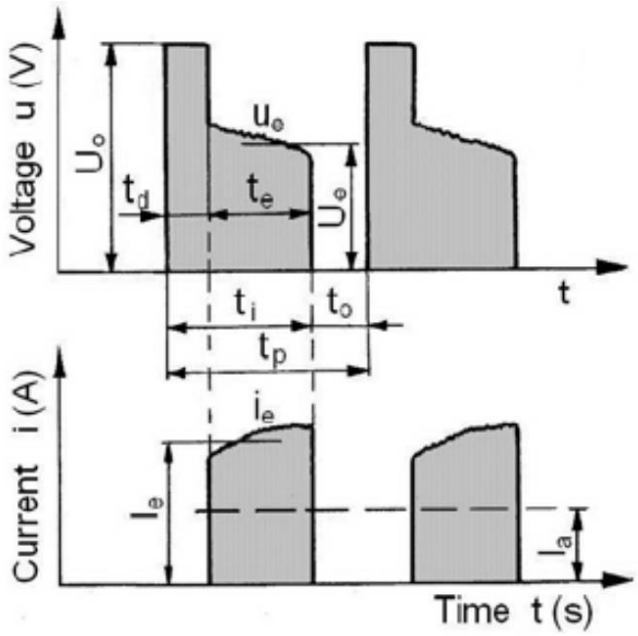
2.3.1. Parametry procesu FHD

W procesie elektroerozyjnego drążenia otworów (FHD), wyszczególnić można szereg podstawowych parametrów procesu. Kombinacja każdego z tych parametrów przekłada się na jakość wykonywanych detali, czas obróbki, niezawodność oraz powtarzalność obróbki. Wiedza na temat parametrów obróbki jest niezwykle istotna, dla zrozumienia procesu. Nieprawidłowe parametry prowadzić mogą do całkowitego zniszczenia detalu. Większość obecnych maszyn FHD ma zdolność do regulacji następujących parametrów procesu.

- Czas impulsu (T_{on}): Czas impulsu jest czasem, w trakcie którego napięcie jest podawane między elektrodą roboczą, a obiektem obrabianym (rysunek 11 – oznaczony jako t_i). W początkowym czasie impulsu zachodzi zjawisko jonizacji dielektryka. Okres przed wytworzeniem iskry nazywany jest czasem opóźnienia zapłonu t_d. Reszta czasu impulsu to okres trwania wyładowania. Generowana iskra topi materiały elektrody i obiektu obrabianego oraz zachodzi odparowanie usuniętego materiału. Dłuższy czas impulsu powoduje usuwanie większej ilości materiału.

- Czas bezimpulsowy (T_{off}): Po czasie impulsu następuje czas bezimpulsowy (rysunek 11 – oznaczony jako t_o). W tym czasie napięcie pomiędzy elektrodami zostaje odcięte przez generator impulsów. Czas bezimpulsowy wykorzystywany jest do przywrócenia właściwości dielektrycznych w płynie oraz pozwala na usunięcia zanieczyszczeń i produktów ubocznych obróbki ze szczeliny iskrowej. Zbyt krótki czas bezimpulsowy może prowadzić do gromadzenia się zanieczyszczeń, co skutkuje niestabilnością obróbki.

- Prąd impulsu (I): Prąd impulsowy to średnia wartość prądu wyładowania w pojedynczym cyklu impulsowym. Maszyny FHD pozwalają zwykle na płynne ustawianie prądu impulsowego.



Rysunek 11 Przebieg charakterystycznych wartości elektrycznych procesu EDD [48]

- Napięcie serwomechanizmu (V): Napięcie serwomechanizmu kontroluje odległość między elektrodą, a obiektem obrabianym, czyli kontroluje wymiar szczeliny iskrowej. Wewnętrzy mechanizm sterowania maszyny utrzymuje stałą odległość szczeliny iskrowej na podstawie kontroli średniego napięcia międzyelektrodowego. Im niższe napięcie serwomechanizmu, tym mniejsza szczelina iskrowa i odwrotnie. W trakcie prowadzenia procesu materiał elektrod jest usuwany powiększając tym samym szczelinę iskrową. Zwiększenie szczeliny przekłada się na zwiększenie napięcia między elektrodami. Układ sterowania maszyny obserwuje zmianę napięcia i przesuwa oś roboczą wrzeciona tak aby zmniejszyć szczelinę i utrzymać napięcie serwomechanizmu na stałym poziomie.

- Ciśnienie płukania (P_f): Ciśnienie płukania kontroluje ciśnienie dielektryka pompowanego przez środek elektrody. W większości maszyn ciśnienie płukania można regulować. Wartość ciśnienia jest wybierana w zależności od wymagań prędkości procesu i wymiarów wierconych otworów. Powiększenie ciśnienia może poprawiać usuwanie zanieczyszczeń, ale zbyt wysokie ciśnienie może powodować niedokładności geometryczne.

- Prędkość obrotowa elektrody (V_z): elektroda w czasie drążenia jest stale wprowadzana w ruch obrotowy w celu jej stabilizacji. Prędkość obrotów może mieć istotny wpływ na proces i zależy od jego parametrów jak np. stosowany typ elektrody.

Stosunek czasów T_{on} i T_{off} nazywany jest często obciążeniem cyklu. Im większy stosunek czasu impulsowego w całym okresie T EDM (rysunek 11 – oznaczony jako t_p), tym większy współczynnik obciążenia.

Na rysunku 11 przedstawiono ideowy przebieg napięcia i prądu procesu EDM [48]. Dokładna lista parametrów może różnić się w zależności od stosowanej maszyny, natomiast przedstawiony powyżej zestaw parametrów jest zwykle uniwersalny niezależnie od konkretnej konstrukcji.

2.3.2. Usterki procesu FHD

Jak opisywano wcześniej, stosowanie procesu FHD jest wymuszone ze względu na charakterystykę obrabianych części. Proces ten niestety jest obarczony pewnymi wyzwaniami w jego kontroli. Obrabiane części są kosztowne, w każdej części wykonywana jest seria nawet kilkuset otworów, gdzie każdy z nich powinien być wykonany prawidłowo, aby część była prawidłowo chłodzona w czasie pracy silnika.

Uproszczony schemat do procesu produkcyjnego części pokazane jest na rysunku 12.

Rysunek 12 Schemat procesu produkcji części wymagających drążenia FHD

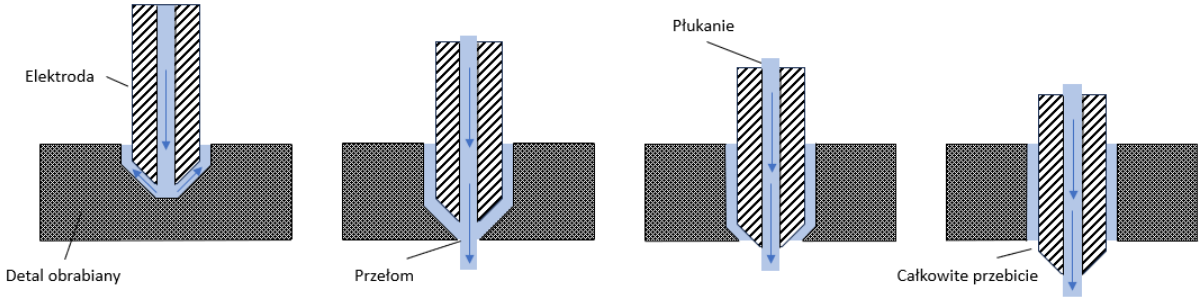
Części obrabiane są najczęściej wytwarzane z superstopów w procesie odlewania kierunkowego. Przed procesem wytwarzania otworów chłodzących często poddawane są innym procesom obróbczym jak wgłębne EDM i szlifowanie. Tak przygotowana część zostaje poddana procesowi FHD. W pojedynczej części drążone jest od kilkudziesięciu do kilkuset otworów. Po drążeniu każda część przechodzi szereg sprawdzeń weryfikujących

poprawność wykonania. Każdy z wywierconych w detalu otworów skanowany jest z pomocą mikroskopu skaningowego 3D w celu weryfikacji poprawności geometrycznej i oceny jakości powierzchni wykonanych otworów. Następnie każdy z otworów poddawany jest testom przepływowym. Przez otwory pompowana jest najpierw woda, a następnie gaz pod ciśnieniem w celu potwierdzenia drożności i charakterystyki przepływowej istotnej z punktu widzenia generacji efektu filmu chłodzącego. Poprawna część następnie zostaje pokryta specjalną powłoką antyoksydacyjną i ponowne sprawdzana. Zdarza się, że w procesie nanoszenia powłok, część z otworów chłodzących traci swoje właściwości, co wymaga poddaniu ich ponownemu procesowi FHD. Na końcu część jest prześwietlana promieniami roentgena, aby potwierdzić ostatecznie jej prawidłowość.

Ze względu na dużą cenę części oraz skomplikowany, pracochłonny i kosztowny proces kontroli produkowanych części, bardzo istotne staje się wprowadzenie metod, które pozwolą na maksymalne ograniczenie defektów produkcyjnych.

Jednymi z podstawowych defektów jakie mogą wystąpić w czasie procesu drążenia otworów są: zbyt wczesne zakończenie drążenia oraz tzw. overdrilling, double-drilling i scarfing.

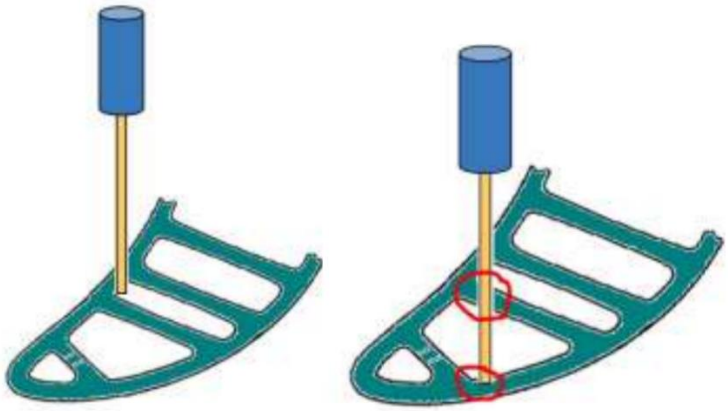
Zbyt wczesne, błędne wykrycie przebicia powoduje zatrzymanie procesu przed całkowitym uformowaniem otworu. Na rysunku 13 pokazano kolejne etapy przebijania elektrody przez detal i stopniowe formowanie zakończania drążonego otworu.



Rysunek 13 Stopniowe formowanie zakończenia otworu FHD

Zakończenie otworu w procesie FHD formowane jest stopniowo. Przed całkowitym przebiciem wierconego otworu dochodzi do fazy tzw. przełomu. W tym momencie tworzony jest już niewielki otwór do wewnętrznego kanału obrabianej części. Proces musi być kontynuowany jeszcze przez jakiś czas, aby stworzyć otwór o jednolitej średnicy.

Zakończenie drążenia zaraz po przełomie owocowało będzie otworem, który będzie zbyt mocno dławił przepływ gazu chłodzącego. Tak wykonany otwór musi zostać poddany ponownemu rozwiercaniu. Generuje to niepotrzebne koszty, ponieważ ten typ defektu może być wykryty dopiero po zakończeniu drążenia wszystkich otworów, na etapie kontroli jakości. Dodatkowo powoduje to ryzyko opisywanego dalej double-drillingu.



Rysunek 14 Wykrywanie przebicia otworowania FHD. Z lewej - poprawne. Z prawej - z błędną detekcją przebicia ((https://www.makino.com/en-us/makino-expertise,13.01.2024)

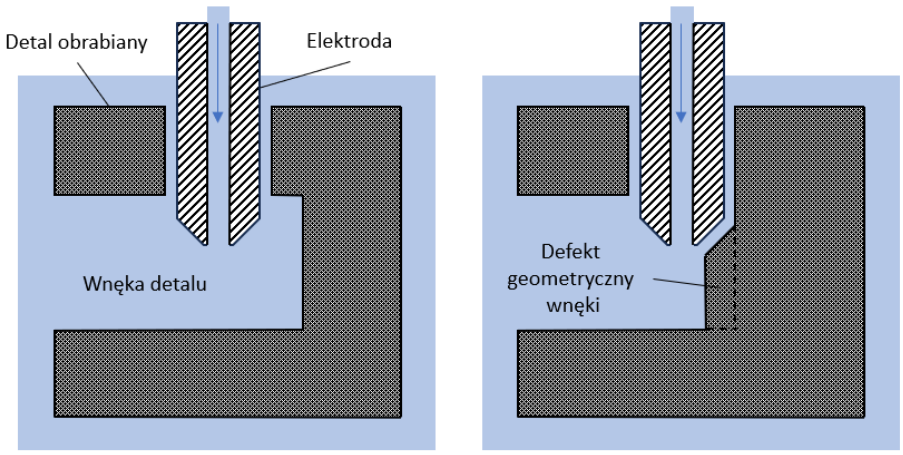
Overdrilling odnosi się do sytuacji, w której po przebiciu ścianki detalu proces nie zostaje zatrzymany odpowiednio szybko. Elektroda po przejściu przez ściankę detalu i dotarciu do wnęki wciąż jest przesuwana w głąb części, gdzie może natrafić na przeciwległą ściankę wnęki i usunąć z niej część materiału. Zjawisko to jest też często nazywane backwall strike. Przykład prawidłowo wykonanego otworu oraz sytuacji z błędnym wykryciem przebicia przedstawiono na rysunku 14.

Ze względu na różne orientacje otworów i stochastyczny charakter procesu FHD (proces o dużej zmienności), wykrywanie całkowitego przebicia jest zadaniem wyjątkowo trudnym. Ze względu na ciągłe zużywanie się elektrody w niejednostajnym tempie nie jest możliwe jednoznaczne powiązanie przesunięcia osi Z wrzeciona obrabiarki z głębokością drążenia. Z tego względu konieczne jest opracowanie innych metod wykrywania przebicia.

Double-drilling jest usterką procesu naprawczego, kiedy konieczne jest rozwiercenie już istniejącego otworu np. w wypadku zbyt wczesnego wykrycia przebicia lub zmniejszenia średnicy otworu w procesie nanoszenia powłok antykorozyjnych. W usterce tej dochodzi do ponownego wiercenia otworu w nieprawidłowym miejscu, przez co zamiast okrągłego otworu powstaje błędny otwór o nieregularnym kształcie. Możliwość wystąpienia tego typu

błędu w czasie drążeń korekcyjnych oraz zwiększony czas produkcji dodatkowo wymusza opracowanie niezawodnych metod kontroli procesu.

Następną formą usterek, które występują na obrabianych detalach jest tzw. scarfing. Taką sytuację schematycznie przedstawiono na rysunku 15.



Rysunek 15 Scarfing w procesie FHD

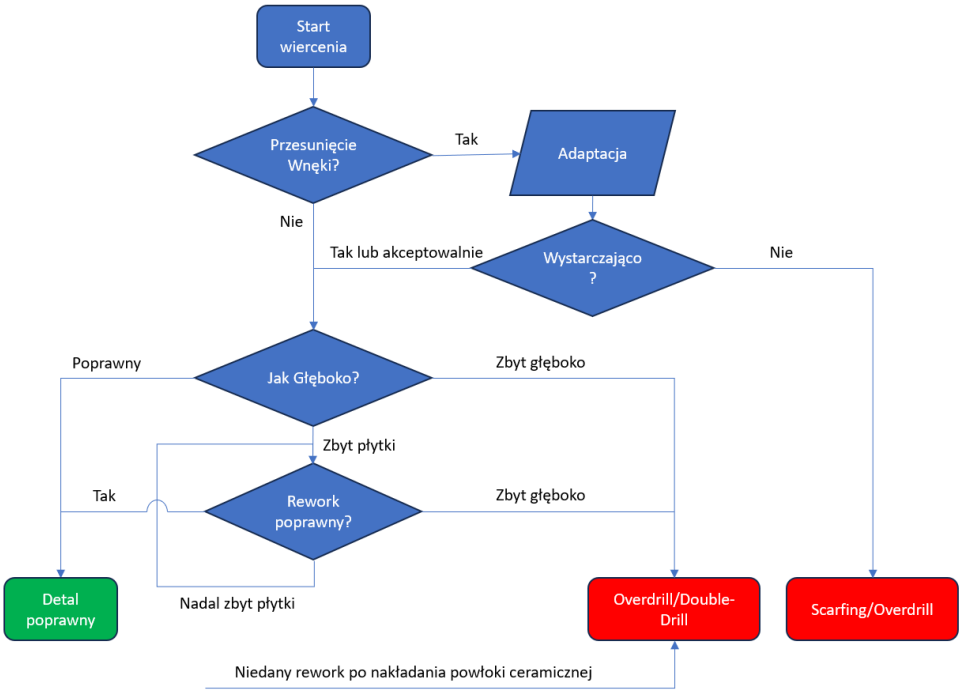
W odlewach poddawanych wierceniu zdarzają się błędy geometryczne położenia wewnętrznego kanału części. Ze względu na zamkniętą i skomplikowaną geometrię części, tego typu niezgodności często nie są wykrywane przed drążeniem otworów. Może to spowodować sytuację, w której pomimo drążenia otworu w prawidłowym położeniu określanym na zewnętrznej powierzchni detalu, elektroda nie wykonuje pełnego okrągłego otworu do wewnętrznego kanału. Zamiast tego następuje dalsze usuwanie materiału częścią średnicy elektrody.

Scarfing jest sytuacją niepożądaną z kilku powodów. Powstały otwór często nie spełnia wymagań przepływowych i negatywnie wpływa na wytrzymałość części. Dodatkowo, wykrycie przebicia standardowymi metodami jest wyjątkowo trudne, ponieważ proces generacji iskier nie jest przerywany, co dla większości normalnych narzędzi kontroli procesu powoduje brak zatrzymania drążenia. Prowadzi to zwykle do over-drillingu i konieczności wykonania napraw lub całkowitego odrzucenia części. Scarfing prowadzi też do nadmiernego zużywania elektrod i wydłużonego czasu obróbki.

Rozwiązanie pozwalające na sprawne wykrywanie przebicia lub zaistnienia scarfingu oraz zatrzymanie procesu w odpowiednim momencie jest niezbędne dla utrzymania wysokiej

jakości produkowanych części oraz zapobiegania generacji dużych kosztów napraw. W wypadku niewykrycia przebicia, detal obrabiany jest całkowicie niezdatny do użytku lub musi zostać poddany kosztownemu procesowi naprawczemu. W wypadku części silników lotniczych, każda niezgodność produkcyjna musi zostać odpowiednio udokumentowana i może zostać dopuszczona do użytku tylko za zgodą inspektora z odpowiednimi uprawnieniami FAA/EASA, co generuje zwykle duże opóźnienia i koszty. Całość procesu drążenia otworu w zależności od grubości ścianki i średnicy elektrody, może trwać zaledwie kilka sekund. Sprawia to, że zastosowane metody muszą wykrywać etapy procesu w czasie rzeczywistym. Opracowane metody pomiaru, przetwarzania danych i wnioskowania na ich podstawie muszą charakteryzować się wystarczającą wydajnością, niezawodnością oraz terminowością tzn. wykrywaniem zaistniałych zjawisk w poprawnym czasie i z niskim opóźnieniem.

Przykładowy diagram prowadzenia procesu wiercenie otworów wentylacyjnych przedstawiono na rysunku 16.



Rysunek 16 Diagram etapów prowadzenia procesu FHD otworów chłodzących

Oprócz zdefiniowanych powyżej błędów procesowych istotne jest także bieżące kontrolowanie przebiegu procesu. Ze względu na stochastyczny charakter procesu, potencjalne problemy niejednorodności materiałów lub np. zużycie lub uszkodzenie elementów maszyny, możliwe jest prowadzenie procesu w środowisku odbiegającym od optymalnych parametrów.

Czynniki te mogą przekładać się na defekty otworów, które będą eliminować je z użytku w faktycznych silnikach. Wyjątkowo trudno jest przewidzieć wszystkie potencjalne przyczyny takich anomalii, a ich symulowanie w warunkach produkcyjnych jest kosztowne i bardzo niepraktyczne. Dlatego zasadne jest opracowanie metod automatycznego kontrolowania i wykrywania anomalii procesu, które pozwolą na wykrywanie potencjalnie wadliwych części już podczas prowadzenia procesu lub po jego zakończeniu, ale bez konieczności przeprowadzania dodatkowych operacji sprawdzających jakość części. Dodatkowo, podejście takie może pozwolić na identyfikowanie i eliminowanie przyczyn powstawania defektów przez optymalizację kolejnych drążeń.

W celu opracowania rozwiązań zapobiegających występowaniu usterek oraz wykrywaniu awarii koniczne jest wykonanie badań eksperymentalnych oraz opracowanie narzędzi pozwalających na lepsze zrozumienie przebiegu procesu oraz usprawniające przetwarzanie danych. Proces tworzenia iskier jest procesem wysokoczęstotliwościowym i charakteryzuje się dużą zmiennością. Fakt ten wymusza rejestrowanie dużej ilości danych do prawidłowego obrazowania zachodzących zjawisk. Próba bezpośredniego przeglądu danych i wyciągania wniosków na tej podstawie byłaby wyjątkowo pracochłonna. Z tego względu konieczne jest zastosowanie zautomatyzowanych metod przetwarzania danych usprawniających ten proces.

W kolejnych rozdziałach opisane zostanie proponowane podejście pozwalające na szybkie i sprawne wykrywanie zaistnienia przebić i anomalii procesu.

3. Przegląd literatury

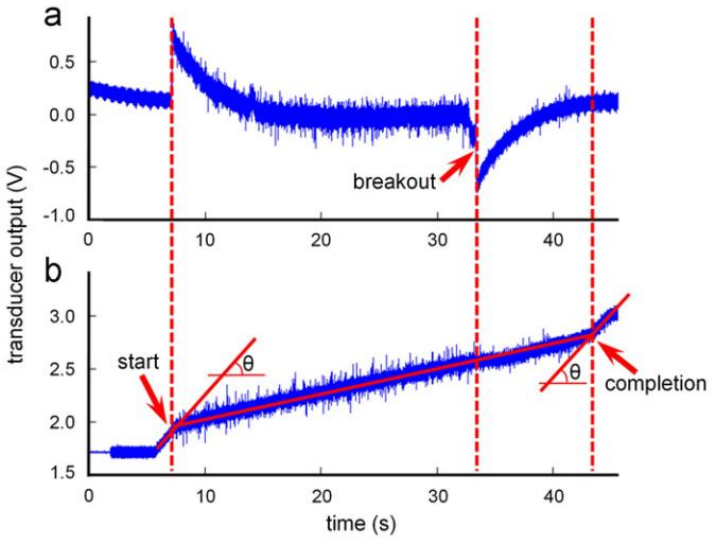
Jak opisano w rozdziale 2.3.2, opracowanie sprawnych narzędzi kontroli procesu FHD dla otworów chłodzących części turbin jest zagadnieniem kluczowym dla prawidłowego prowadzenia procesu w warunkach produkcyjnych. Badacze zaproponowali różne podejścia do zagadnienia wykrywania przebicia i przebiegu analizy procesu. W rozdziale 3.1 opisane zostały dotychczasowe postępy prac naukowych w tej dziedzinie. Jak pokazano w toku analizy literaturowej, dotychczasowe prace w niewielkim stopniu korzystały z zalet jakie do zagadnienia wprowadzać mogą zaawansowane metody uczenia maszynowego, dlatego w kolejnych częściach przeglądu literaturowego zawarto przegląd metod ML i DL stosowanych do monitorowania i kontroli procesów przemysłowych (rozdziały od 3.3 do 3.6) oraz przeanalizowano podejścia do wykrywania anomalii (rozdziały od 3.5 do 3.9)

3.1. Wykrywanie przebicia elektrody w procesie FHD

Istnieje wiele prac opisujących na zagadnienia dotyczące procesu EDM, analizowane było w nich wiele różnych aspektów procesu, między innymi w dziedzinach, charakteryzowania procesu [49-52], optymalizacji obróbki [44, 53, 54], modelowania procesu [55-58], czy wykrywania przebicia [59-69]. Jednym z głównych zagadnień niniejszej pracy jest monitorowanie przebiegu procesu oraz opracowanie metod i narzędzi do pewnego wykrywania pełnego przebicia elektrody, czyli pełnego zakończenia formowania otworu w wierconym elemencie. W niniejszym rozdziale zawarto przegląd dotychczasowych prac badawczych podejmujących próbę opracowania takich podejść i rozwiązań.

Do tej pory zaproponowano kilka metod detekcji zakończenia drążenia otworu, głównie w przemyśle. B. Sciaroni [59] zamocował metalową płytę zatrzymującą pod przedmiotem obróbki, która może generować sygnały kontaktowe, gdy elektroda przewierci przedmiot obrabiany i dotrze do płyty. Oczywiście ta metoda nie nadaje się do detali z kanałami wewnętrznymi. Znaczna większość części lotniczych obrabianych w procesie FHD posiada skomplikowane kanały wewnętrzne, co eliminuje możliwość stosowania tej metody. Dodatkowo, zanieczyszczony odłamkami płyn roboczy może być przewodzący i generować fałszywe sygnały. K.B. Haefner et al. [60] sugerowali, że wiercenie powinno być natychmiast zatrzymywane, gdy częstotliwość wyładowań elektrycznych nagle wzrasta, co wskazuje,

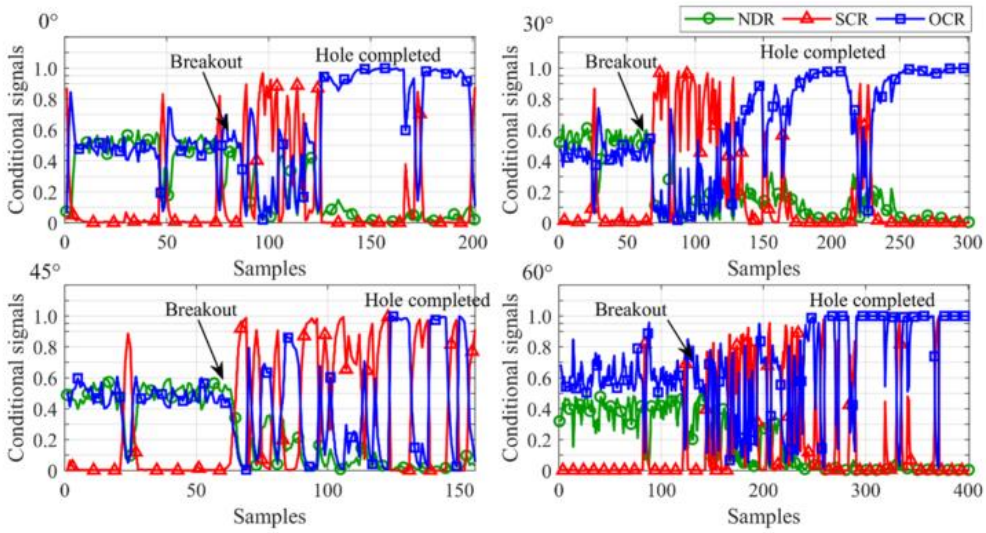
że elektroda już dotarła na drugą stronę kanału. W podejściu tym w sposób kontrolowany dopuszcza się do powstania błędu wiercenia typu „overdrilling”. Ta metoda zasadniczo nie wykrywa etapu przebicia, a jedynie ogranicza szkody poprzez ograniczenie głębokości na jaką elektroda może zagłębić się w przeciwległej ściance wewnętrznego kanału części obrabianej. Yamada et al. [61] zaprojektowali układ logiczny do przetwarzania sygnałów, takich jak aktualne napięcie szczeliny, przepływ i ciśnienie płynu roboczego, głębokość obróbki oraz kierunek ruchu elektrody w celu generowania sygnałów wskazujących, czy elektroda przewierciła przedmiot obróbki czy nie. Pewnym ograniczeniem podejścia był fakt, że układ logiczny mógł jedynie wykonywać proste operacje binarne. Z tego powodu jest on nieprecyzyjny w detekcji zakończenia otworu. Koshy et al. [62] zalecali, aby zakończenie drążenia otworu było uznane wtedy, gdy średnia prędkość posuwu elektrody równa się prędkości posuwu jaka była na początku procesu. Badacze ponadto, opierali metodologię wykrywania przełomu na monitorowaniu ciśnienia w płynie roboczym. Dowodzili, że wystąpienia przełomu powinno powodować nagły spadek ciśnienia płukania elektrody (Rysunek 17).



Rysunek 17 Przebiegi ciśnienia płukania (a) i sygnału przemieszczenia wrzeciona w procesie FHD [62]

Ta metoda jest prawdopodobnie niepraktyczna, ponieważ elektroda jest często cofana w fazie przełomu, zwłaszcza podczas wiercenia otworów pod kątem np. ze względu na zaistnienie zwarć. Metoda ta także nie bierze pod uwagę możliwość występowania wyładowań po zaistnieniu przełomu ze względu na pogorszenie warunków płukania. Bellotti et al. [63] sugerowali, że monitorowanie średniej częstotliwości normalnych wyładowań jest niezawodnym podejściem do określania zakończenia otworu. Jednak nie przedstawili

dalszych dyskusji ani wyników badań eksperymentalnych. Metody detekcji przełomu były również badane przez Lin i Nien [64]. Wykrywali oni przełom, analizując widmo sygnału napięcia wyładowania i prędkość posuwu elektrody. Badania W. Xia et al. [65, 66] skupiały się również na detekcji przebicia w oparciu o sygnały napięcia i prądu elektrod oraz prędkość posuwu elektrody narzędziowej. Identyfikowali oni relatywną ilość różnych typów wyładowań i ich związek z etapem procesu. Zaproponowali metodę opartą o klasyfikację grafów stanu obróbki (CMSG). Graf stanu obróbki (MSG) był formowany na podstawie zmieniających się cech wzorcowych sygnałów, które zmieniają się gwałtownie podczas wystąpienia przełomu. Następnie problem detekcji rozwiązano poprzez klasyfikację MSG w czasie rzeczywistym. Jednak prace skupiały się tylko na wykrywaniu początkowej fazy przełomu i metoda nie dostarczała wystarczających informacji do detekcji zakończenia otworu. Nadal potrzebna jest skuteczna i praktyczna metoda detekcji zakończenia otworu. Kontynuację prac nad tym podejściem przedstawili Zhang et al. [67], gdzie zaproponowali nową metodę detekcji zakończenia otworu poprzez analizę sygnałów wyładowczych w fazie przebicia. Metodologia wykrywania przebić skupiała się na analizie charakteru wyładowań, poprzez wprowadzenie zmiennych warunkowych (Rysunek 18): stosunek wyładowań normalnych (NDR), stosunek zwarć (SCR) i stosunek obwodu otwartego (OCR). W badaniu przypisanie każdego z wyładowań do odpowiedniej grupy wykonywane było na podstawie prostego schematu decyzyjnego, w którym klasyfikacja była przeprowadzana poprzez porównanie próbkowanych wartości prądu i napięcia z ustalonymi na stałe wartościami progowymi.



Rysunek 18 Zmiany sygnałów warunkowych podczas wiercenia normalnego otworu (67)

Badacze dowodzą, że pomimo obecności każdego z typów wyładowań w czasie wiercenia, w fazie przebicia zachodzą wspólne tendencje zmian tych sygnałów. W podejściu zaproponowano identyfikację momentu zakończenia formowania otworu, kiedy relatywna częstotliwość występowania obwodów otwartych (OCR) znacznie wzrasta, a występowanie zwarć (SCR) i wyładowań normalnych spada do wartości bliskiej zeru. Bellotti et al. [68] badali sygnał napięcia szczeliny w procesie wiercenia otworów mikroskopijnych przy użyciu EDM i zaproponowali podejście oparte na charakterystyce procesu wyładowczego do detekcji przełomu. Opierali swoje podejście na ocenie odstępów czasowych między kolejnymi nietypowymi wyładowaniami. W tym samym duchu zastosowali metodę sterowania sumą kumulatywną do detekcji zakończenia otworu poprzez monitorowanie zmian w stosunku normalnych wyładowań. Wang et al. [69] opracowali metodę identyfikacji etapu przebicia opartą na kurtozie (KBSI). Metoda ta wykorzystuje znormalizowaną kurtozę do określenia zmian trendów sygnałów napięcia szczeliny, umożliwiając identyfikację etapu procesu w czasie trwania drążenia. Badacze dowodzą, że znormalizowana kurtoza może przedstawiać zmienność napięcia wyładowania w szczelinie. Proponują metodę, w której identyfikacja zakończenia etapu przełomu następuje w chwili, gdy sygnał znormalizowanej kurtozy spada ponownie do małych wartości.

Jak pokazuje przegląd literatury, obecny trend badań wykrywania przebicia i ogólnej kontroli procesu w dużej mierze opiera się na charakteryzacji czasowych sygnałów prądowych i napięciowych elektrod. Aby lepiej zrozumieć te podejścia i móc prawidłowo interpretować zbierane dane, konieczny jest przegląd zagadnienia klasyfikacji wyładowań EDM. Kolejny rozdział przedstawia obecny stan wiedzy na temat tego zagadnienia.

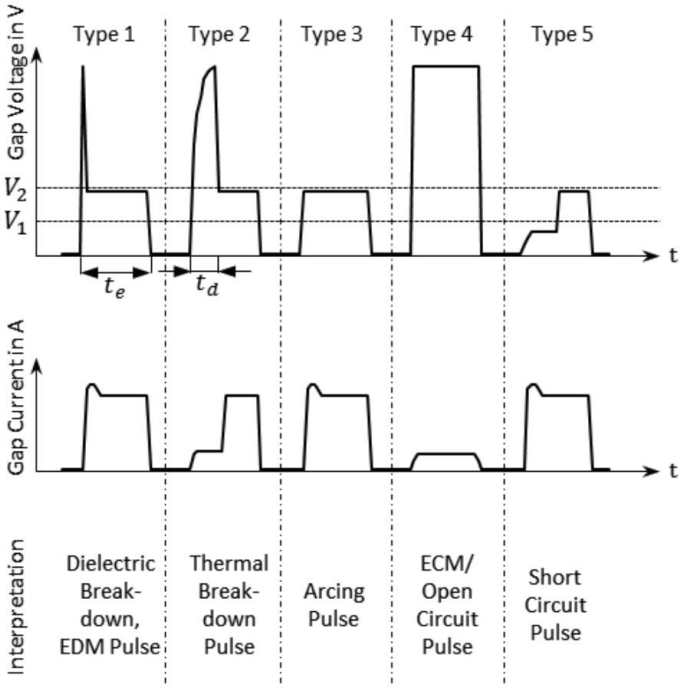
3.2. Analiza przebiegu prądu i napięcia procesu EDM

Jak wykazał przegląd literatury, większość nowoczesnych podejść stosowanych do identyfikowania etapów procesu FHD opiera się o analizę wysokoczęstotliwościowych charakterystyk przebiegów prądu i napięcia elektrod. Niestety, podczas analizy literaturowej nie zidentyfikowano źródeł odnoszących się specyficznie do procesu wiercenia otworów w technologii EDM. Założyć można natomiast, że sam proces generacji iskier powinien mieć zbliżoną charakterystykę do wyładowań generowanych podczas drążeń wgłębnym EDM czy WEDM.

W procesie elektrodrążenia (EDM), charakterystyczne parametry sygnałów napięcia i prądu posiadają standardowe kształty [49-52]. Iskry wytwarzane mogą być z częstotliwością cyklu generatora EDM złożonego z czasu impulsu T_{on} i czasu bezimpulsowego T_{off}. Jednak, ze względu na charakter procesu, nie każdy z okresów generuje wystąpienie normalnej iskry. W literaturze impulsy w procesach elektrodrążenia klasyfikowane są zwykle do czterech głównych typów: obwód otwarty, normalne wyładowanie, łuk i zwarcie. W niektórych publikacjach wyładowania normalne są jeszcze dzielone na podkategorie ze względu na sposób ich tworzenia i przebieg [52].

Impulsy normalne spowodowane przełamaniem dielektrycznym (Rysunek 19, Type 1) wykazują charakterystyczne przebiegi napięcia i prądu wyładowania iskrowego o niskim czasie opóźnienia zapłonu t_d. Te impulsy mogą być generowane w wąskiej szczelinie drążenia. Ten typ impulsu intensywnie zużywa elektrodę, ale charakteryzuje się dobrą szybkością usuwania, co przyśpiesza drążenie poszczególnych otworów.

Impulsy normalne termiczne (Rysunek 19, Type 2), w których czas zapłonu jest opóźniony. Podczas czasu opóźnienia zapłonu t_d, obserwuje się znaczny zwolniony przyrost napięcia z towarzyszącym mu przepływem prądu wstępnego. Prąd wstępny powoduje lokalne ogrzewanie dielektryka w wyniku efektu Joule'a, co doprowadza ostatecznie do przełamania spowodowanego wysoką gęstością prądu. Impulsy te zwiększają szybkość erodowania materiału, a tym samym produktywność.



Rysunek 19 Schematyczne kształty sygnałów prądu i napięcia w elektrodrążeniu [52]

Wyładowania łukowe (Rysunek 19, Type 3), nie wykazują czasu opóźnienia zapłonu, a napięcie nie wzrasta powyżej napięcia wyładowczego. Maksimum prądu jest podobne do prądów wyładowań typu 1 i 2, co wskazuje na wyładowania łukowe. Wyładowania łukowe przyczyniają się do wysokich szybkości usuwania materiału, ale także zwiększają chropowatość powierzchni obrabianego przedmiotu.

Impulsy ECM lub impulsy obwodu otwartego (Rysunek 19, Type 4), charakteryzują się małym prądem roboczym i napięciem pomiędzy napięciem iskrowym, a napięciem obwodu otwartego. Czynnik roboczy jest podgrzewany, a usuwanie elektrochemiczne jest wywoływane działaniem elektrolitycznym płynu dielektrycznego. Amplituda prądu zależy od średniej przewodności elektrycznej elektrolitu i jego składników gazowych. Zaletą tego typu impulsu jest niskie zużycie elektrody oraz zwykle osiąganie dobrej jakości powierzchni.

Impulsy zwarciowe (Rysunek 19, Type 5), można podzielić na kolejne podkategorie. Tzw. „miękkie” zwarcie, podobne do impulsów termicznych, w których napięcie obwodu otwartego nie jest osiągane, co jest możliwe przy bardzo dużym zanieczyszczeniu szczeliny. Drugim typem zwarć są impulsy „twardego” zwarcia spowodowane bezpośrednim kontaktem elektrod. Impulsy „miękkiego” zwarcia zazwyczaj występują tylko okresowo i mogą być usuwane poprzez płukanie. Impulsy zwarcia powinny być unikane, ponieważ nie przyczyniają się do usuwania materiału oraz znacząco obniżają jakość powierzchni.

Każdy z typów wyładowań może występować w czasie procesu drążenia otworów z różnym prawdopodobieństwem w zależności od parametrów obróbki oraz etapu procesu. Zrozumienie typów impulsów jest kluczowe w celu prawidłowego wykrywania przebicia. Odpowiednia klasyfikacja i wykrywanie typów impulsów może także pomóc w wykrywaniu anomalii procesu w celu wykrywania potencjalnych defektów i optymalizacji prowadzenia procesu.

3.3. Inteligentna diagnostyka w monitorowaniu procesów technologicznych

Niniejsza rozprawa skupia się na opracowaniu rozwiązań dla optymalizacji procesu produkcji otworów chłodzących części turbin silników odrzutowych. Ze względu na stopień skomplikowania procesu i jego stochastyczny charakter opracowane rozwiązania powinny mieć możliwość prawidłowego rozpoznawania i interpretowania potencjalnie niebezpiecznych dla procesu obróbczego zjawisk bez konieczność interakcji z człowiekiem. W zasadzie oznacza to konieczność opracowania opartego na danych inteligentnego systemu diagnostyki usterek i monitorowania przebiegu procesu produkcyjnego.

Inteligentna diagnoza usterek odnosi się do zastosowań tradycyjnych technik uczenia maszynowego, takich jak sztuczne sieci neuronowe (ANN), maszyny wektorów nośnych (SVM), drzew decyzyjnych oraz nowocześniejszych metod opartych o uczenie głębokie z wykorzystaniem głębokich sieci neuronowych (DNN), do diagnozowania usterek maszyn [70, 71]. Metody te posiadają potencjał do osiągania celów diagnostycznych przy ograniczeniu zaangażowania ludzi. Takie metody wykorzystują teorie uczenia maszynowego do adaptacyjnego uczenia się wiedzy diagnostycznej maszyn na podstawie zebranych danych, zamiast korzystać z doświadczenia i wiedzy inżynierów. Konkretnie inteligentne systemy diagnostyczne mają na celu konstrukcję modeli diagnostycznych zdolnych do automatycznego łączenia relacji między zebranymi danymi, a stanem maszyn.

Wiele prac naukowych jest przedmiotem badania różnych metod uczenia maszynowego jako podstawy dla inteligentnych systemów diagnostycznych. Dotyczy wielu gałęzi produkcji i metod obróbkowych jak np. kontrola prawidłowości przebiegu procesu produkcyjnego i wykrywanie awarii komponentów maszyn. Pomimo oczywistych różnic wynikających z samych zjawisk fizycznych występujących w czasie prowadzenia danego procesu, zagadnienie opracowania inteligentnego systemu diagnostycznego dla procesu FHD nadal będzie mieć

wiele czynników wspólnych i borykać się będzie z podobnymi problemami jak inne metody produkcyjne. Dlatego w niniejszym rozdziale dokonano przeglądu literatury dotyczącej aktualnego stanu wiedzy w dziedzinie rozwiązań inteligentnych systemów diagnostycznych i zastosowanych w nich metod ML.

3.3.1. Diagnostyka z wykorzystaniem uczenia maszynowego

Tradycyjne teorie uczenia maszynowego były z sukcesem stosowane w inteligentnych systemach diagnostycznych. Wykorzystanie tradycyjnych metod, takich jak algorytm k-najbliższych sąsiadów (kNN) [72], ANN [73], SVM [74], i probabilistyczny model graficzny (PGM) [75] znacznie poszerzyło możliwości wielu dziedzin techniki. Stosowanie podejść opartych na ANN [76], podejść opartych na SVM [77] i innych inteligentnych podejść [70] przyczyniło się do powstania inteligentnych systemów diagnostycznych. W tych metodach cechy usterek były wyodrębniane z zebranych danych procesowych. Proces wymagał wybrania cech istotnych z punktu widzenia szkolenia modeli diagnostycznych, które na ich podstawie automatycznie określały stan maszyny.

Stosując tradycyjne metody ML konieczne jest przeprowadzenie zbierania danych, generowanie cech i wnioskowanie na ich podstawie. Duża część pracy polega na ręcznym badaniu stanów pracy maszyn i procesów oraz określaniu istotnych cech, co zwiększa intensywność pracy i obniża dokładność diagnozy. Poniżej przedstawiono przykładowe podejścia stosowane do tej pory w każdym z etapów opracowania rozwiązań inteligentnych systemów diagnostycznych.

Inteligentna diagnostyka z zastosowaniem ML - zbieranie danych

W celu zbierania danych wykorzystuje się odpowiednio dobrane czujniki, które są zamontowane na maszynach i w sposób ciągły mierzą wybrane wielkości charakteryzujące proces. Zazwyczaj stosuje się różne sensory, mierzące wiele wielkości fizycznych takich jak np. prąd, napięcie, temperaturę, wibracje, emisję akustyczną. Zazwyczaj zbieranie danych odbywa się jednocześnie z różnych źródeł przy użyciu różnorodnych czujników i systemów wizji maszynowej [78]. Prowadzone badania udowodniły, że stosowanie danych z wielu źródeł pozwalają na osiągnięcie lepszej dokładności diagnozy niż przy użyciu danych z pojedynczego czujnika ze względu na obecność uzupełniających się informacji [79].

Inteligentna diagnostyka z zastosowaniem ML – przetwarzanie danych

W celu poprawy wydajności systemów monitorujących, przed zastosowaniem modeli uczenia maszynowego wymagana jest odpowiednia obróbka sygnałów. Powszechnie stosowane metody przetwarzania sygnałów prowadzą do wyodrębnienia cech z dziedziny czasu, z dziedziny częstotliwości lub z połączenia dziedziny czasu i częstotliwości.

Cechy w dziedzinie czasu mogą być wymiarowe i bezwymiarowe. Cechy wymiarowe to np. wartość średnia, odchylenie standardowe, średnia kwadratowa, wartość szczytowa, itp. Do stosowanych cech bezwymiarowych możemy zaliczyć kurtozę, wskaźnik kształtu, skośność, wskaźnik grzbietu, wskaźnik luzu, wskaźnik impulsu, itp. [80, 81].

W dziedzinie częstotliwości stosowane są cechy takie jak średnia częstotliwość, środek częstotliwości, odchylenie standardowe częstotliwości, itp. [80, 81]. Stosowanie zmiennych z dziedziny częstotliwości często pozwala na uzyskanie informacji, których nie da się określić tylko na podstawie cech z dziedziny czasu.

Z połączenia czasu i częstotliwości wyodrębnić można cechy takie jak np. entropia energii [80, 81]. Cechy tego typu zazwyczaj tworzone są za pomocą transformacji falkowej (WT), transformacji pakietu falkowego (WPT) lub dekompozycji empirycznej modelu. Te cechy są w stanie odzwierciedlić stan maszyn w warunkach niestabilnego działania.

Inteligentna diagnostyka z zastosowaniem ML – wybór cech

Aby opracować inteligentny model podejmowania decyzji konieczne jest wybranie optymalnych cech procesu, ponieważ część z wyodrębnionych cech może posiadać redundantne dane o procesie [82, 83]. Zbyt duża ilość cech zwiększa koszt obliczeniowy oraz może prowadzić do tzn. przekleństwa wymiarowości [84]. Najczęściej stosowane metody wyboru cech obejmują wrappery, filtry oraz metody wbudowane [85].

Metody oparte na wrapperze [86] skupiają się na ocenie istotności cech na podstawie wydajności trenowanych klasyfikatorów. Wybierany jest podzbiór cech i sprawdza się z ich pomocą efektywność klasyfikacji. Czynność jest powtarzana dla kolejnych podzbiorów cech, aż klasyfikator uzyska najkorzystniejszą wydajność. Popularnym rozwiązaniem jest algorytm Las Vegas Wrapper (LVW) [87].

Metody oparte na filtrach przetwarzają zebrane cechy, niezależnie od użytego klasyfikatora [86]. Jako przykłady filtrów wyszczególnić można np. metodę zysku informacyjnego i stosunku zysku [88], określanie wskaźnika istotności, oceniającego wrażliwość cech na stan diagnozowanego procesu [89], wybór cech najmniej podobnych metodą Minimalnej Redundancji Maksymalnego Związku (mRMR) [90] czy selekcja na podstawie odległości, przez jednoczesną minimalizację odległości wewnątrzklasowej i maksymalizację odległości międzyklasowej [91, 92].

Metody wbudowane integrują selekcję cech w proces trenowania klasyfikatorów. Automatycznie wybierają cechy po zakończeniu trenowania narzucając warunki regularyzacji na obiekty optymalizacji [86]. Zazwyczaj stosowane są regularyzacja L1 [93] lub regularyzacja L2 [94]. Regularyzacja może złagodzić problem przeuczenia, który występuje podczas trenowania na małej liczbie próbek.

Niemniej jednak, pomimo skuteczności metod optymalizujących wybór cech, wiele badań prowadzi ekstrakcję cech metodą prób i błędów lub odwołuje się do podobnej literatury bez przeprowadzania dogłębnych badań nad optymalnym wyborem cech.

Inteligentna diagnostyka z zastosowaniem ML – wnioskowanie

Rozpoznawanie stanu procesu wykorzystuje modele diagnozowania oparte na uczeniu maszynowym do ustanowienia związku między wybranymi cechami, a określonymi stanami procesu. Model oparty na ML potrzebuje najpierw zostać nauczony z pomocą oznakowanych danych wejściowych. Przygotowane w ten sposób modele są w stanie rozpoznać stany maszyny, gdy wejściowe próbki są nieoznaczone.

W kolejnej sekcji przedstawione zostały najpopularniejsze rozwiązania stosowane jako modele decyzyjne w inteligentnych systemach diagnostycznych.

Metoda k najbliższych sąsiadów kNN

Metoda k najbliższych sąsiadów kNN (ang. k-Nearest Neighbors) to powszechnie używany model klasyfikacyjny oparty o uczenie nadzorowane [72]. W tej metodzie miara odległości jest używana do wyszukiwania k próbek w pobliżu danej nieoznaczonej próbki.

kNN jest szeroko stosowany w badaniach dotyczących inteligentnych systemów diagnostycznych, zwłaszcza w diagnozie usterek silników [95-97] łożysk tocznych [98, 99] czy kół zębatych [100 - 102]. Jednak kNN ma pewne problemy, takie jak trudność w wyborze optymalnej wartości parametru sąsiedztwa i nieokreśloną granicę sąsiedztwa. Niektórzy badacze zbadali specjalnie zmodyfikowany kNN i użyli ich do diagnozy usterek maszyn. Zhao et al. [103] zaproponowali euklidesowy klasyfikator ważony kNN do monitorowania usterek łożysk, gdzie odległość euklidesowa była używana do ważenia cech w celu podkreślenia ich znaczenia w klasyfikacji. Dong et al. [104] użyli algorytmu optymalizacji rojem cząstek do optymalizacji kNN i uzyskali wyższą dokładność diagnozy dla łożysk niż inne metody bez optymalizacji.

Modele wnioskowania oparte na kNN są łatwe do wdrożenia, ale wymagają dużej mocy obliczeniowej. Parametr k jest trudny do ustalenia, co znacznie utrudnia tworzenie optymalnych modeli. Nierównomierny rozkład danych diagnostycznych może obniżać dokładność wnioskowania tego rodzaju modeli.

Metody oparte na sztucznych sieciach neuronowych (ANN)

ANN próbuje swoim działaniem imitować ludzki mózg w przetwarzaniu informacji, co stanowi skuteczną metodę budowy modeli diagnozy [73, 106].

W literaturze znaleźć można wiele przykładów zastosowania sztucznych sieci neuronowych (ANN), do kontroli stanu silników elektrycznych i spalinowych, łożysk tocznych oraz przekładni. Pod względem metodologii, BPNN, falowa sieć neuronowa (WNN) i radialna sieć funkcji bazowych (RBFN) są szeroko stosowane do realizacji zadań diagnozy. Publikacje często opisują specjalne modyfikacje wprowadzone do podstawowych ANN w celu polepszenia ich wydajności. Merainani et al. [107] użyli samoorganizującej się mapy cech w sieci neuronowej, aby identyfikować stany zdrowia automatycznych skrzyń biegów w różnych trybach pracy. Chen et al. [108] wykorzystali probabilistyczną sieć neuronową do efektywnej diagnozy usterek jednostek generatorów hydraulicznych. Shifat et al. [109] z sukcesem zaproponowali rozwiązanie do wykrywania usterek silników BLDC korzystając z wielu czujników. Barakat et al. [110] wprowadzili rozwijającą się sieć neuronową do konstrukcji modelu diagnozy łożysk silnikowych, który osiągnął wyższą dokładność diagnozy dla dużej liczby danych w porównaniu z konwencjonalną RBFN i probabilistyczną siecią neuronową. Chikkam et al. [111] stosowali sztuczne sieci neuronowe do klasyfikacji

usterek silników indukcyjnych przez analizę sygnałów silników poddanych dyskretnej analizie falkowej.

ANN posiadają pewne wady. Modele diagnozy oparte na ANN są czarnymi skrzynkami [112] ze względu na brak rygorystycznego wsparcia teoretycznego. W rezultacie są one trudne w interpretowaniu. Dodatkowo, złożoność modeli diagnozy znacznie wzrasta wraz z zwiększeniem ilości monitorowanych danych wejściowych. Rosnąca ilość parametrów modelu obniża efektywność szkolenia i prowadzi do nadmiernego dopasowania, co obniża dokładność diagnozy modeli. Mimo tego dzięki wysokiej zdolności do samouczenia, modele diagnozy oparte na ANN mogą automatycznie uczyć się wiedzy diagnostycznej z danych wejściowych, minimalizując ryzyko empiryczne. Ponadto, mogą łatwo rozpoznawać wiele stanów maszyn.

Metody oparte na SVM (Support Vector Machine)

SVM [74] to metoda stosowana często w zadaniach klasyfikacji. Jest jednym z bardzo popularnych algorytmów opartych na uczeniu nadzorowanym. SVM rozróżnia dwie klasy, znajdując optymalną hiperpłaszczyznę, która maksymalizuje margines między najbliższymi punktami danych przeciwnych klas.

SVM pełni rolę powszechnie stosowanej metody uczenia maszynowego w zastosowaniach inteligentnych systemów diagnostycznych, zwłaszcza w diagnozowaniu usterek przekładni, łożysk, silników i układów wirujących [113, 114] oraz urządzeń hydraulicznych [115]. Dla tych zadań klasyfikacji, modele oparte na SVM rozpoznają zwykle wiele stanów, a nie tylko stan poprawnej pracy i stan awarii. Prace rozwojowe nad SVM skupiają się głównie na odpowiednim zmodyfikowaniu i optymalizacji algorytmu. Przykładami użycia zmodyfikowanych SVM zastosowanego do diagnozowania usterek maszyn mogą być: SVM zbliżeniowe [116], SVM z najmniejszymi kwadratami [117, 118], SVM jednej klasy [119, 120], SVM falkowy [121-123], SVM zespołowe [124, 125], SVM rozmyte [126, 127], SVM z wieloma jądrami [128, 129]. Podejścia te zwykle osiągały lepszą wydajność klasyfikacji niż konwencjonalne SVM. Część badaczy wzbogacało SVM o algorytmy optymalizacyjne, takie jak algorytm genetyczny [130,131], optymalizacja kolonii mrówek [132] i optymalizacja rojem cząsteczek [133-135].

Modele oparte o SVM odznaczają się pewnymi problemami. Algorytm SVM został pierwotnie stworzony do rozwiązywania zadań klasyfikacji binarnej. W przypadku zadań wieloklasowych w diagnostyce stanu procesu konieczne jest użycie skomplikowanych podejść, takich jak OAA (one-against-all) i OAO OAA (one-against-one), do zintegrowania wyników z wielu modeli opartych na SVM. Dodatkowo takie modele są najskuteczniejsze dla małej liczby danych wejściowych. Zwykle są one trudne do implementacji dla dużej ilości danych, co prowadzi do dużej złożoności obliczeniowej. Skuteczność klasyfikacji SVM jest wrażliwa na parametry jądra. Nieodpowiednie parametry mogą całkowicie uniemożliwić uzyskanie poprawnych wyników. Pomimo swoich wad modele oparte na SVM są szkolone na podstawie minimalizacji ryzyka strukturalnego, co przyczynia się do poprawy interpretowalności. Rozwiązanie optymalizacyjne SVM odnosi się do wyznaczania globalnego minimum w strukturalnej optymalizacji kwadratowej, przez co dla części zagadnień klasyfikatory mogą łatwo osiągnąć wysoką dokładność diagnozy i uzyskać optimum globalne.

Drzewa decyzyjne

Drzewa decyzyjne to kolejna popularna metoda uczenia nadzorowanego w zagadnieniach klasyfikacji. Algorytm ustala związek między klasą, a cechami za pomocą struktur w kształcie drzewa.

Wśród wielu metod drzew decyzyjnych, algorytm C4.5 jest skuteczny w uzyskiwaniu satysfakcjonującej dokładności oraz posiada zrozumiałe reguły klasyfikacji [136]. Drzewo decyzyjne typu C4.5 zostało skutecznie zastosowane do diagnozy usterek układów rotorowych [137], łożysk tocznych [138], pomp odśrodkowych [139] i skrzyń biegów [140]. Pewną ewolucją podejścia drzew losowych jest las losowy [141], który łączy decyzje generowane z wielu klasyfikatorów. Podejście to poprawia możliwości generalizacji drzewa decyzyjnego. Podejście lasu losowego zostało też wdrożone w zagadnieniach diagnostycznych. Li et al. [142] zastosowali metodę lasu losowego do budowy klasyfikatorów w skrzyniach biegów, co pozwoliło im osiągnąć wysoką poprawność wyników. Tun et al. [144] zaproponowali hybrydową budowę lasów losowych do klasyfikacji usterek systemów HVAC. Wang et al. [143] diagnozowali usterki łożysk tocznych przez modele diagnostyczne oparte na lasach losowych. Wiele prac prowadzone jest także w kierunku poprawy wydajności lasów losowych. Dla przykładu Asadi et al. [145] użyli algorytmu optymalizacji roju cząstek do optymalizacji parametrów lasu losowego.

Pewnym minusem modeli w kształcie drzewa jest fakt, że są one w większości konstruowane na podstawie wiedzy ekspertów. Tego typu modele łatwo ulegają przeuczeniu i mają małą zdolność do generalizacji. Pomimo tych minusów drzewa decyzyjne charakteryzują się także pewnymi zaletami. Modele kontroli zbudowane na drzewach decyzyjnych są łatwe w interpretowaniu, przez co można je łatwo przekształcić w reguły diagnozy. Dość dobrze radzą sobie także z brakującymi danymi wejściowymi.

Probabilistyczne modele graficzne (PGM)

Probabilistyczne modele graficzne są modelami probabilistycznymi do obrazowania związków między danymi za pomocą struktur graficznych. Węzły reprezentują zbiory zmiennych, a połączenia między nimi oznaczają probabilistyczny związek między danymi.

Najczęściej stosowane klasyfikatory są oparte na modelach bayesowskich i Markova. Yu et al. [146] zastosowali naiwny klasyfikator bayesowski i elastyczny naiwny klasyfikator bayesowski do wykrywania uszkodzeń kół zębatych. Wang et al. [147] użyli klasyfikatora bayesowskiego do identyfikacji usterek systemów energetycznych. Yu et al. [148] używali klasyfikatora bayesowskiego do monitorowania zdrowia kół zębatych. Naiwny klasyfikator bayesowski podlega założeniu, że pomiędzy cechami nie występują zależności. Aby uniknąć tego ograniczenia nienaiwny klasyfikator bayesowski [149] został opracowany. Podejście to zostało zastosowane do diagnozy usterek np. w pracy Yu et al. [150] do diagnozowania łożysk tocznych. Modele Markova zostały użyte do klasyfikacji stanu pomp hydraulicznych [151] oraz łożysk [152]. Część publikacji podjęło próbę poprawy wydajności modeli Markova. Huang et al. [153] używali sieci neuronowej i zbiorów rozmytych do budowania macierzy obserwacji ukrytych modeli Markova. Poprawiło to dokładność diagnozy stanu systemów napędowych. Xiao et al. [154] przedstawili model oparty na szkoleniu wielu ukrytych modeli Markova do diagnozy usterek łożysk. Pozwoliło to na połączenie informacji z wielu kanałów wejściowych.

Modele PGM są trudne do implementacji dla złożonych związków ze względu na niską zdolność dopasowywania danych. Wymagają też wcześniejszego poznania probabilistycznych związków między zmiennymi, co często nie jest oczywiste. Pomimo tego, stosując podejście PGM, stosunkowo łatwo uzyskać klasyfikację usterek z wieloma stanami procesu badanego. Dodatkowo PGM mogą z powodzeniem być używane do analizy różnic między klasami.

3.3.2. Diagnostyka z wykorzystaniem głębokiego uczenia

Teorie głębokiego uczenia zaczęły reformować środowisko produkcyjne w tym inteligentne systemy diagnostyczne począwszy od 2010 [155].

Uczenie głębokie [73, 155] podąża za ideą uczenia efektywnych reprezentacji z danych poprzez szkolenie elastycznych, wielowarstwowych („głębokich") sieci neuronowych i znacznie poprawiło stan techniki w wielu aplikacjach, które obejmują złożone typy danych. Głębokie sieci neuronowe dostarczają najbardziej udanych rozwiązań dla wielu zadań w dziedzinach takich jak widzenie komputerowe [156, 157], rozwój nauki [158, 159], rozpoznawanie mowy [160-162] czy przetwarzanie języka naturalnego [163, 164]. Metody oparte na głębokich sieciach neuronowych potrafią wykorzystać hierarchiczną lub ukrytą strukturę, która często występuje w danych, poprzez ich wielowarstwowe, rozproszone reprezentacje cech. Postępy w obliczeniach równoległych, optymalizacji SGD i automatycznym różniczkowaniu umożliwiają zastosowanie uczenia głębokiego na dużą skalę przy użyciu dużych zestawów danych.

Chociaż systemy diagnostyczne w przeszłości potrafiły samodzielnie rozpoznawać stany badanego procesu, w wielu zastosowaniach analiza dużych lub wysokowymiarowych danych za pomocą konwencjonalnych metod ML wymagała nadal dużego zaangażowania ludzkiego w proces wyodrębniania cech. Głębokie uczenie to nowy temat w dziedzinie uczenia maszynowego, którego główną zaletą jest uniknięcie tej przypadłości metod ML. Podczas, gdy termin „uczenie głębokie" odnosi się do stosowania licznych warstw ukrytych w strukturze ANN [13], różni się głównie od tradycyjnych modeli ML tym, w jaki sposób modele są uczone na podstawie surowych danych. Podstawową zaletą DL w porównaniu do tradycyjnego ML, jest możliwość prowadzenia inżynierii cech w strukturze modelu podczas procesu uczenia. Zdolności wysokopoziomowej reprezentacji abstrakcyjnej i inżynierii cech sprawiają, że modele DL są odporne na wariacje danych [13]. Ponadto, hierarchiczna struktura głębokich sieci pozwala im modelować złożone nieliniowe relacje w dużych danych.

Teorie głębokiego uczenia dalej inspirowały rozwój inteligentnych systemów diagnostycznych i doprowadziły do szeregu osiągnięć [13, 20-22], w tym podejść opartych na stosowanych autoenkoderach (AE), podejść opartych na sieciach głębokich przekonań DBN, konwolucyjnych sieciach neuronowych (CNN), rekurencyjnych sieciach neuronowych

(RNN) i podejść wykorzystujących ewolucję tych metod jak np. ResNet. W tych metodach DL pozwala na automatyczne uczenie się stanów procesu na podstawie surowych zebranych danych, zamiast sztucznego wyodrębniania cech. Modele te mogą automatycznie ustanawiać związek między pomiarami, a docelowym wynikiem, bezpośrednio łącząc surowe dane monitorowania z odpowiadającymi im stanami badanego procesu. Pozwala to na dalsze ograniczenie wkładu pracy ludzkiej.

Modele diagnostyczne stosowane w inteligentnych systemach diagnostycznych składają się zwykle z warstwy ekstrakcji cech i warstwy klasyfikacji. Modele najpierw korzystają z sieci hierarchicznych, aby uczyć się stopniowo abstrahowanych cech. Następnie, warstwa wyjściowa umieszczona po ostatniej warstwie ekstrakcji cech rozpoznaje stan procesu, zazwyczaj przy użyciu klasyfikatora opartego na ANN ze względu na jego dużą zdolność do klasyfikacji wieloklasowej. W trakcie procesu szkoleniowego błąd między prawdziwym stanem, a przewidzianym przez system jest minimalizowany za pomocą algorytmu wstecznej propagacji błędu, który aktualizuje parametry treningowe modeli diagnozy. W dalszym rozdziale opisane zostaną najczęściej stosowane metody głębokiego uczenia i ich zastosowania w diagnozie stanu procesu produkcyjnego.

Metody oparte na Autoenkoderach (AE)

Autoenkodery (AE) to nienadzorowane, jednokierunkowe sieci neuronowe, w których wyjście stara się odwzorować dane wejściowe. Składają się warstw enkodera i dekodera, w których pierwsza przesyła dane wejściowe do warstwy ukrytej, a druga odtwarza dane wejściowe z tej reprezentacji [165]. Zwykle, stosowane są algorytmy oparte na gradientach do dostrojenia hiperparametrów modelu poprzez minimalizację błędu rekonstrukcji. Główne warianty AE to autoenkodery z zakłóceniem i rzadkie autoenkodery (SAE) [13].

Niektóre publikacje przedstawiły AE i jego powszechne odmiany w diagnozie usterek maszyn. AE został zastosowane do diagnozy usterek maszyn w pracy Jia et al. [166] gdzie użyto wielu AE do automatycznego uczenia cech z danych dziedziny częstotliwości. Skonstruowany model diagnozy zawierał trzy ułożone AE, które pomagały automatycznie separować nieużyteczne informacje i kompresować pomocne informacje, zamiast ręcznego ekstrahowania cech statystycznych. Wyniki pokazały, że zaproponowana metoda miała potencjał radzenia sobie z ogromnymi danymi monitorującymi i uzyskiwania wysokiej dokładności diagnozy.

Lu et al. [167] użyli ułożonych AE usuwających zakłócenia do diagnozy usterek łożysk, a wyniki badań wykazały wyższą dokładność diagnozy niż inne metody, takie jak SVM i ANN. Dou et al. [168] zastosowali AE do monitorowania zużycia narzędzi przy użyciu sygnałów wibracji i siły w procesie frezowania. Bezpośrednio przekazywali segmenty sygnałów do modelu SAE i pokazali, że w miarę wzrostu zużycia narzędzi błąd rekonstrukcji stawał się bardziej niestabilny. Mogli zidentyfikować cztery stany zużycia narzędzi przy użyciu proponowanego modelu monitorowania. Ocha et al. [169] użyli stosowanych, rzadkich autoenkoderów do klasyfikacji zużycia narzędzi w procesie frezowania aluminium przy użyciu czujników siły, wibracji i emisji akustycznej. Kim et al. [170] również użyli opartej na progach metody AE do rozróżniania nowego i zużytym narzędzia, korzystając z czujników siły skrawania i prądu.

Aby poprawić wydajność modeli diagnozy opartych na AE, badacze proponowali wiele podejść optymalizacyjnych. Na przykład Liu et al. [171] użyli AE do budowy rekurencyjnej sieci neuronowej do diagnozy usterek łożysk silnika. Shi et al. [172] zaproponowali konwolucyjne autoenkodery uczenia ekstremalnego, w których parametry są losowo ustalane, a nie przy użyciu algorytmu propagacji wstecznej. Dalej sieci uczone były nieoznaczonymi danymi w postaci hiper-grafów do wykrywania usterek maszyn rotujących. Innym przykładem zastosowania autoenkoderów uczonych ekstremalnie jest praca Mao et al. [173]. Shao et al. [173] zaproponowali użycie funkcji falowej Gaussa jako funkcji aktywacji do zaprojektowania falowego AE. Zbudowany na tej podstawie głęboki autoenkoder falowy pozwolił na poprawę wyników kontroli stanu łożysk tocznych lokomotyw. Ponadto badacze starali się rozwijać hybrydowe modele diagnozy, łącząc AE z innymi metodami. Liu et al. [174] użyli splotowych autoenkoderów 1D w celu usunięcia zakłóceń z danych oraz konwolucyjnych sieci neuronowych 1D do klasyfikacji stanu obrabiarek. DBN to inna metoda do budowy hybrydowych modeli diagnozy z użyciem ułożonych AE [175, 176]. W modelach diagnozy ułożone AE odpowiedzialne są za modelowanie cech z danych wejściowych, podczas gdy DBN służy do rozpoznawania stanów zdrowia maszyn na podstawie nauczonych cech.

Jak wykazał przegląd literatury, autoenkodery jako metoda uczenia nienadzorowanego, nie mogą być bezpośrednio używane do rozpoznawania stanów zdrowia maszyn. Dlatego w inteligentnych systemach diagnostycznych dodaje się zwykle warstwę klasyfikacji na szczycie architektury modelu, a skonstruowane modele diagnozy muszą być szkolone na wystarczającej liczbie oznaczonych próbek. Modele diagnozy oparte na stosowych AE

są w stanie automatycznie reprezentować informacje o stanie maszyn z danych wejściowych oraz nie wymagają dużego wkładu pracy ekspertów w proces ekstrakcji cech.

Modele oparte na DBN

Sieć głębokich przekonań (DBN) składa się ułożonych sekwencyjnie ograniczonych maszyn Boltzmanna (RBM) [177]. W DBN istnieją połączenia między warstwami, ale nie między neuronami w obrębie jednej warstwy. Struktura warstwowo-warstwowa sieci zapewnia hierarchiczną reprezentację cech [178], która jest używana do wysokopoziomowego reprezentowania danych wejściowych.

W trakcie procesu treningu nienadzorowanego, DBN odbudowuje swoje dane wejściowe poprzez naukę rozkładu prawdopodobieństwa. RBM jest efektywnym narzędziem do inżynierii cech. Trening sieci DBN obejmuje trening wielu RBM, gdzie ukryta warstwa niższego RBM jest uważana za dane treningowe modelu, a wyjście RBM jest używane jako dane treningowe dla wyższego RBM. Po przeszkoleniu wszystkich RBM-ów, proces dostrojenia jest wykonywany poprzez zastosowanie algorytmu wstecznej propagacji z danymi treningowymi jako wyjściem [179, 180].

DBN okazała się skuteczną metodą w badaniach dotyczących inteligentnych systemów diagnostycznych. Na przykład w pracy Tamilselvan et al. [181] DBN zostało zastosowane do diagnozowania usterek silników lotniczych, co było jednym z pierwszych badań w tej dziedzinie. Jiang et al. [182] stosowali wiele RBM do konstrukcji modeli diagnozowania opartych na DBN, które uzyskiwały wyższą dokładność diagnozy niż tradycyjne metody diagnozy uszkodzeń łożysk tocznych. Podobne zastosowanie miała praca He et al. [183] gdzie zaprezentowano model diagnozowania oparty na pojedynczym Gaussowskim RBM. Zhu et al. [184] zbadali model klasyfikacji uszkodzeń łożysk z pomocą DBN gdzie danymi uczącymi były zredukowane z pomocą PCA sygnały wibracji. Yu et al. [185] zaproponowali model diagnozowania usterek sterowany danymi dla turbin wiatrowych, który również był realizowany przez DBN. W części z badań diagnozowania usterek urządzeń hydraulicznych [186] i silników indukcyjnych [187] DBN osiągało wyższą dokładnością niż tradycyjne metody.

Aby poprawić wydajność diagnozy, badacze zbadali dalej algorytm optymalizacji dla modeli diagnozowania opartych na DBN. Niu et al. [188] wprowadzili optymalizację

parametrycznej funkcji aktywacji ReLU przy użyciu roju cząstek. Shao et al. [189] próbowali także adaptacyjnie określić strukturę modeli diagnozowania opartych na DBN, używając algorytmu roju cząsteczek. He et al. [190] użyli DBN do diagnozowania usterek w łańcuchu przekładniowym przekładni zębatej, a algorytm genetyczny został dodatkowo użyty do zoptymalizowania struktury DBN. Jin et al. [191] testowali system kontroli łożysk przekładni osi pociągu z wykorzystaniem DBN. Stosowali wariacyjną dekompozycję trybów (VMD) na sygnałach wibracji oraz metodę Grey Wolf (GWO) do optymalizacji parametrów dekompozycji.

W przeciwieństwie do ułożonych AE, modele diagnozowania oparte na DBN mogą automatycznie uczyć się cech z danych wejściowych, przetwarzając zestaw ułożonych RBM, co rozwiązuje problem zanikającego gradientu przy użyciu algorytmu wstecznej propagacji (BP) do dostrojenia sieci o głębokich warstwach. Aby rozpoznać stany zdrowia maszyn, DBN mapuje nauczane cechy do przestrzeni etykiet, dodając warstwę klasyfikacji. Konieczne jest używanie wystarczającej ilości opisanych danych do szkolenia skonstruowanych modeli diagnozowania, aby uzyskać przekonujące wyniki diagnozy.

Metody oparte na CNN

Konwolucyjna sieć neuronowa (CNN) to zregularyzowany typ sieci neuronowej typu feed-forward, która samodzielnie uczy się inżynierii cech poprzez optymalizację filtrów (lub jąder). Zazwyczaj, CNN składa się z warstw koewolucyjnych, warstw poolingowych i warstw w pełni połączonych. Ukryta warstwa sieci obejmuje szereg warstw konwolucyjnych, które abstrahują tensor wejściowy do mapy cech, za pomocą wielu lokalnych filtrów jądra. Filtry są konwoluowane wzdłuż szerokości i wysokości danych wejściowych, generując dwuwymiarowe mapy cech [192]. Otrzymaną mapę cech można połączyć z warstwą w pełni połączoną w celu przeprowadzenia nadzorowanej regresji lub klasyfikacji.

CNN jako nadzorowana metoda uczenia maszynowego, odniosła kilka znakomitych osiągnięć w rozpoznawaniu mowy, identyfikacji obrazów i śledzeniu celów [193].

Architektury CNN można podzielić na modele diagnozy oparte na 2-wymiarowych (2D) CNN oraz modele diagnozy oparte na 1-wymiarowych (1D) CNN. Pierwotnie, CNN 2D opracowane zostały jako narzędzie do identyfikacji obrazów. Dane wejściowe dla tych modeli są danymi 2D. W przypadku diagnozy awarii maszyn, jednakże 2D CNN

nie jest w stanie obsłużyć sygnałów 1D, takich jak dane wibracyjne. Badacze zmodyfikowali standardowe podejście CNN 2D tak, aby radziło sobie także z danymi jednowymiarowymi. Stworzony 1D CNN zaczyna radzić sobie z danymi wibracyjnymi, które podlegają cechom zmienności przesunięcia, i skutecznie pomógł skonstruować modele diagnozy dla przekładni [194-196], łożysk tocznych [197, 198] i silników [199, 200], gdzie dane wejściowe do modeli diagnozy stanowiły surowe dane, bez przetwarzania wstępnego.

Aby skonstruować model diagnozy oparty na CNN 2D i uzyskać wysoką wydajność, badacze zastosowali kilka skutecznych podejść, które można podzielić na trzy typy:

1. Stosowanie danych obrazowych, takich jak obrazy odcieni szarości i obrazy termowizyjne [201, 202], były stosowane w diagnozie awarii maszyn. W rezultacie modele diagnozy oparte na CNN mogą bezpośrednio działać skutecznie, podobnie jak w przypadku zadań klasyfikacji identyfikacji obrazów.

2. Metody przetwarzania sygnałów, takie jak pakiety falkowe [203, 204], ciągła transformata falkowa [205, 206] i transformata synchrosqueezing [208], są wykorzystywane do przetwarzania wstępnego wejściowych danych 1D, które mają na celu przekształcenie sygnałów z dziedziny czasu na dziedzinę czasu-częstotliwość. Następnie CNN może obsługiwać dane monitorowania z widmem czasowo-częstotliwościowym w formie 2D.

3. Niektóre publikacje [209] ręcznie zmieniają wymiary danych wejściowych, aby dostosować je do modeli diagnozy opartych na CNN. W tych pracach, przetwarzanie wstępne danych realizowane jest głównie z pomocą prostych i pomysłowych operacji, a nie zaawansowanych metod przetwarzania sygnałów, co niemal eliminuje konieczność korzystania z wiedzy ekspertów.

W porównaniu z modelami opartymi na stosowanych AE i DBN, modele diagnozy oparte na CNN są w stanie bezpośrednio uczyć się cech z surowych danych monitorowania bez przetwarzania wstępnego, takiego jak transformacja częstotliwościowa, ponieważ CNN potrafi uwzględnić właściwości zmienności przesunięcia danych wejściowych. Ponadto liczba parametrów treningowych w modelach diagnozy jest zmniejszana przez dzielenie wag, co może przyspieszyć zbieżność i powstrzymać nadmierne dopasowanie. Podobnie

jak w przypadku innych modeli opartych na głębokim uczeniu, wydajność diagnozy modeli opartych na CNN zależy od trenowania ich z dostateczną ilością oznaczonych próbek.

Metody oparte na RNN

Sieć neuronowa rekurencyjna (RNN) to klasa jednokierunkowych sztucznych sieci neuronowych (ANN), zdolna do aktualizacji bieżącego stanu na podstawie bieżących danych wejściowych oraz poprzednich stanów. Dlatego też jest idealna do pracy z danymi sekwencyjnymi, szeregami czasowymi lub sygnałami niepodzielonymi, przechwytując informacje przechowywane sekwencyjnie w poprzednich elementach. RNN korzystają z uczenia nadzorowanego, które cierpi z powodu znikającego gradientu. Dlatego standardowe sieci neuronowe rekurencyjne (RNN) nie są w stanie efektywnie radzić sobie z długotrwałymi zależnościami w danych i są trudne do zastosowania, gdy przerwa w danych wejściowych jest duża [210]. Opracowano różne warianty RNN, spośród których długa pamięć krótkotrwała (LSTM) została szeroko wykorzystana do rozwiązania wspomnianego wyżej problemu.

Badacze zwracają coraz większą uwagę na wykorzystywanie sieci rekurencyjnych (RNN), w szczególności modeli LSTM, do monitorowania stanu narzędzi w trakcie procesów obróbkowych. Ostatnio LSTM zostało użyte do prognozowania chropowatości powierzchni na podstawie detekcji drgań w procesie skrawania [212]. Zou et al. [213] użyli dekompozycji EEMD w celu uzyskania czystego sygnału cech awarii i sieci LSTM do automatycznej ekstrakcji cech awarii. Zhao et al. [214] pokazali wysoką skuteczność sieci LSTM w diagnozowaniu awarii silników. LSTM z sukcesem stosowane był także w wykrywaniu awarii procesu chemicznego [215]. Wykorzystana w badaniach sieć była uczona surowymi danymi procesu bez konieczności ekstrakcji cech. Wyniki badań pokazały dobrą zdolność sieci do rozróżniania wielu klas awarii. Podobnie, Lei et al. [2019] zastosowali LSTM w diagnozie awarii turbin wiatrowych. Pokazali zdolność LSTM do wykrywania długoterminowych zależności i powtarzalnych zjawisk mających wpływ na klasyfikację usterki. Pojawia się także wiele propozycji hybrydowych podejść łączących LSTM z innymi metodami klasyfikacji jak SVM [217-219]. Standardowe sieci RNN także znalazły zastosowanie w systemach diagnostycznych. Liu et al. [220] zastosowali klasyczny RNN w kombinacji z autoenkoderami odszumiającymi do diagnozy stanu łożysk tocznych.

Podejście LSTM ma kilka wad. Trening i ocena LSTMs mogą być kosztowne obliczeniowo, zwłaszcza dla długich sekwencji lub dużych zbiorów danych. Wymagają dużej

ilości danych treningowych do nauki złożonych wzorców w danych. Jeśli dostępne jest niewystarczająco dużo danych, model może nie osiągnąć dobrej wydajności. LSTM mogą być podatne na przeuczenie na małych zbiorach danych, zwłaszcza jeśli architektura modelu jest zbyt skomplikowana. Pomimo swoich wad LSTMs są dobrze przystosowane do modelowania długotrwałych zależności w danych sekwencyjnych, ponieważ mogą selektywnie „pamiętać" lub „zapominać" informacje w czasie. Ponieważ LSTM utrzymują wektor stanu komórki, który reprezentuje „pamięć" sieci w każdym kroku czasowym, mogą być bardziej interpretowalne niż inne rodzaje rekurencyjnych sieci neuronowych. LSTM są bardziej odporne na zakłócenia lub brakujące dane niż inne rodzaje rekurencyjnych sieci neuronowych, ponieważ mogą selektywnie eliminować nieistotne lub szumne informacje za pomocą bramki zapominającej. Te cechy sprawiają, że LSTM ma duży potencjał przy budowaniu inteligentnych systemów diagnostycznych.

Podejścia oparte o ResNet

ResNet jest kolejną ewolucją podejścia do sieci neuronowych realizowaną poprzez dodanie połączonych kilku bloków rezydualnych [221]. Bloki rezydualne składają się z kanału przedniego (forward channel) i połączenia skrótu (shortcut connection).

Badacze przeprowadzili wiele badań nad zastosowaniami ResNet. Model diagnostyczny oparty na analizie czasowo-częstotliwościowej i ResNet został zaproponowany przez Ma et al. [222] do diagnozowania usterek skrzyń biegów planetarnych. Zhang et al. [223] skonstruowali model diagnostyczny oparty na ResNet do diagnozowania łożysk tocznych, w którym surowe dane wibracyjne były bezpośrednio używane do treningu modelu diagnostycznego. Wyniki porównawcze wykazały wyższą dokładność diagnozy w porównaniu do modeli diagnozy opartych na CNN. Hou et al. [224] także zastosowali sieć ResNet w połączeniu z enkoderem transformującym dane wejściowe oraz ideą uczenia transferowego do diagnozowania stanu łożysk. Liang et al. [225] użyli ResNet w diagnozie usterek łożysk. Zaproponowali użycie transformaty falkowej oraz zmodyfikowanej warstwy poolingu sieci, aby przeciwdziałać zakłóceniom sygnału. Dowodzili w swojej pracy, że ich podejście ma większą odporność na zakłócenia od standardowych podejść. Zhao et al. [226] opracowali dynamicznie ważone współczynniki falki dla poprawy wydajności modeli diagnozy opartych na ResNet, uzyskując wyższą dokładność w diagnozie usterki skrzyń biegów planetarnych w warunkach dużych zakłóceń niż inne metody oparte na głębokim uczeniu. Peng et al. [227] użyli ResNet

do rozpoznawania stanów zdrowia łożysk komponentów pociągów. Su et al. [228] skonstruowali model diagnostyczny oparty na ResNet do diagnozowania usterek wózków pociągów wysokiej prędkości. Model wykorzystywał nieprzetworzone szeregi czasowe danych pomiarowych jako dane wejściowe sieci. Wyniki eksperymentów opisanych w publikacji pokazały przewagę modelu diagnostycznego opartego na ResNet nad innymi podejściami głębokiego uczenia się.

ResNet został opracowany w celu uzyskania wyższej wydajności generalizacji w oparciu o architekturę CNN. Dlatego modele diagnozy oparte na ResNet dziedziczą zalety modeli diagnozy opartych na CNN i mogą osiągnąć wysoką dokładność diagnozy, zwłaszcza w warunkach złożonych, takich jak zmienne warunki prędkości lub zmienne warunki obciążenia. Sieci o dużej liczbie warstw (nawet tysiące) mogą być łatwo trenowane bez zwiększania procentowego błędu treningowego.

3.4. Podsumowanie przeglądu literatury o monitorowaniu procesu przebicia FHD

Przegląd literatury pokazuje, że do tej pory zaproponowano wiele metod wykrywania przebicia. Standardowe podejścia podobne do prac Yamady [61] i Koshyego [62] są często w różnych formach implementowane na obecnie dostępnych maszynach FHD. Nie zapewniają one niestety dostatecznej niezawodności. Istnieje potrzeba na opracowanie bardziej skutecznych i niezwodnych metod, co stanowi główny powód przygotowania niniejszej pracy.

Jak widać, duża część z prac badawczych próbuje interpretować etapy procesu na podstawie zarejestrowanych sygnałów napięcia i prądu elektrod, czasami w połączeniu z innymi pomiarami jak np. ciśnienie płukania lub przemieszczenie wrzeciona w osi Z. Często sygnały elektryczne rejestrowane są w dużą częstotliwością pozwalającą na identyfikację pojedynczych impulsów. Jednakże, wspomniane wyżej badania w dużej mierze polegają na wiedzy empirycznej do ustalania progów, które są używane do rozróżniania impulsów wyładowczych lub do odróżniania statusów obróbki poprzez porównanie amplitud lub zmian jednego lub więcej sygnałów. Gdy warunki procesu, takie jak czas trwania impulsu, odstęp między impulsami i prąd szczytowy się zmieniają, progi tych metod stają się mniej ogólne i wymagają dostrojenia. Metody te nie podejmują próby określenia charakteru poszczególnych impulsów. Takie metody nie są naprawdę skuteczne i niezawodne.

W literaturze brakuje opisów dokładnego przebiegu procesu oraz charakteru impulsów występujących konkretnie w procesie FHD. W niniejszej pracy podjęta zostanie próba bardziej szczegółowej analizy wysokoczęstotliwościowych przebiegów wartości elektrycznych odpowiedzialnych za generowanie iskier EDM. Dokonana zostanie identyfikacja typowych kształtów impulsów, ich rozkładu oraz zmian ich proporcji podczas występowania kluczowych etapów wiercenia, jakimi są uformowanie przełomu oraz zakończenie otworu.

Obecne podejścia nie korzystają także z potencjału metod uczenia maszynowego i uczenia głębokiego. Jedyne zastosowanie ML wykorzystane było do tej pory do klasyfikacji przetworzonych wcześniej danych częstotliwości występowania wyładowań w analizie offline, po zakończeniu całego procesu. Zastosowanie metod ML i DL ma potencjał do osiągniecia celów monitorowania procesu bez konieczności dużego nakładu pracy ludzkiej. Dokładna charakteryzacja całości procesu drążenia standardowymi metodami byłaby niezmiernie pracochłonna i niepraktyczna. Podejście to wymagałoby bardzo dużego zaangażowania ekspertów. Metody uczenia i klasyfikacji mogą pozwolić na osiągnięcie wykrywania przebicia w czasie rzeczywistym i wykrywania anomalii bez pracochłonnego przygotowywania danych. Metody DL mogą też być odporne na dużą zmienność charakteru każdego z impulsów, jeśli proces uczenia zostanie przeprowadzony w sposób gwarantujący odpowiednią zdolność do uogólniania wnioskowania. Dokładna klasyfikacja pojedynczych impulsów może dostarczać dodatkowych informacji o przebiegu procesu, ustalenie wpływu ustawianych parametrów na efekty obróbki i potencjalnie może ułatwić korelację występowania defektów z czynnikami wpływającymi na proces.

Niestety do tej pory bardzo mało badań w pełni wykorzystuje metody ML i DL w procesie monitorowania FHD. Dlatego w kolejnym rozdziałach dokonany zostanie przegląd podejść budowy systemów monitorowania procesów produkcyjnych (rozdział 3.3) i wykrywania anomalii (rozdział 3.5) z wykorzystaniem metod ML i DL. Przegląd ma na celu rozpoznanie stanu wiedzy, który umożliwi wybór podejść do przetestowania w części eksperymentalnej pracy. Pozwoli to na dobór optymalnego podejścia do opracowania narzędzi monitorowania procesu FHD, będących przedmiotem niniejszej rozprawy.

Jak pokazuje literatura, tradycyjne techniki uczenia maszynowego mogą w pewnym stopniu być wykorzystane w inteligentnych systemach diagnostycznych. Tego typu podejście może dostarczać zadawalających wyników, natomiast posiada dwie zasadnicze wady.

Wydajność i zdolność samo-nauki tradycyjnych modeli diagnozy są często niewystarczające do ustalenia związku między bardzo dużą ilością zbieranych danych, a odpowiadającymi im stanami maszyn. Dodatkowo, etap sztucznego wyodrębniania cech jest bardzo pracochłonny i jego wyniki zależą od pracy człowieka. Konieczne jest projektowanie potężnych algorytmów do wyodrębniania wrażliwych cech stanów procesu. Często podejście to jest nierealne nawet dla doświadczonych ekspertów z powodu ogromnych kosztów pracy. Dlatego praktycznym podejściem jest opracowanie modeli diagnozy, które będą w stanie automatycznie rozpoznawać stany procesu przy jednoczesnym wyodrębnieniu istotnych cech z surowych zebranych danych wejściowych. Opisywane w kolejnych rozdziałach podejścia głębokiego uczenia mogą dostarczać narzędzi do ominięcia opisywanych wyżej ograniczeń standardowych podejść ML.

Inteligentne systemy diagnostyczne oparte o metodologię uczenia głębokiego pozwalają skonstruować modele end-to-end, które mogą bezpośrednio łączyć relację między surowymi danymi monitorującymi, a stanami procesu lub zdrowia maszyn. Chociaż osiągnięto pewne sukcesy, są one głównie uzależnione od powszechnego założenia: dostępność danych oznaczonych i zawierających pełne informacje na temat stanów zdrowia maszyn [70]. Podczas prowadzenia badań, należy położyć nacisk na zbieranie danych, aby zawierały wystarczająco dużo informacji do odzwierciedlenia poszczególnych rodzajów stanów procesu. Większość zebranych danych nie jest oznaczona. W wypadku wielu procesów nie jest realistyczne częste zatrzymywanie maszyn i sprawdzanie stanów zdrowia ze względu na ogromną utratę zasobów. Co za tym idzie, należy położyć duży nacisk na opracowanie odpowiedniej metodologii przetwarzania danych i automatyzacji opisywania danych. Pomimo tych ograniczeń możliwość znacznego zmniejszenia pracochłonności na etapie przygotowywania danych uczących, raportowane lepsze dokładności klasyfikacji i potencjalna zdolność do radzenia sobie z zakłóceniami sprawiają, że podejścia DL mogą być efektywną metodologią budowania systemów diagnozy także dla procesu EDM.

Bardzo mało prac badawczych analizujących proces EDM korzysta z narzędzi ML i DL, znaczna część prac koncentruje się wokół innych technologii obróbczych jak obróbka skrawaniem czy analiza sygnałów wibracji odwzorowujących stan zdrowia maszyn. Założeniem pracy jest przeniesienie doświadczeń tych prac badawczych w środowisko analizy procesu FHD. Opisywane powyżej zidentyfikowane zalety podejść DL i ML powinny w podobnym stopniu usprawniać proces analizy i diagnozowania procesu EDM. W niniejszej

pracy zostanie zaproponowane podejście do identyfikacji etapów drążenia FHD i wykrywania przebicia oraz narzędzia do klasyfikacji poszczególnych wyładowań EDM oparte o taką metodologię. Przetestowane zostaną różne podejścia oraz przeprowadzone zostanie porównanie wydajności metod DL względem tradycyjnych podejść ML.

3.5. Wykrywanie anomalii

Anomalia to obserwacja, która znacząco odbiega od pewnego pojęcia normalności, znana również jako nowość lub wartość odstająca. Taka obserwacja może być określana jako nietypowa, nieregularna, atypowa, niekonsekwentna, niespodziewana, rzadka, błędna, wadliwa, oszukańcza, złośliwa, nienaturalna lub po prostu dziwna — w zależności od sytuacji. Detekcja anomalii (czyli detekcja wartości odstających lub detekcja nowości) to obszar badawczy zajmujący się analizą wykrywania takich anomalii za pomocą metod, modeli i algorytmów opartych na danych.

Klasyczne podejścia do detekcji wartości odstających obejmują algorytmy PCA [229, 230], OC-SVM [231], SVDD [232], algorytmy najbliższego sąsiada [233] oraz KDE. Są to metody nienadzorowane, co stanowi dominujące podejście w detekcji wartości odstających. Wynika to z tego, że w standardowych ustawieniach detekcji wartości odstających, oznaczone dane odstające często nie istnieją. Jeśli są dostępne, zazwyczaj są niewystarczające do pełnego scharakteryzowania wszystkich pojęć wartości odstających. To zazwyczaj sprawia, że podejście nadzorowane staje się nieefektywne. Zamiast tego, centralnym pomysłem w detekcji wartości odstających jest nauka modelu normalności z danych normalnych w sposób nienadzorowany, dzięki czemu anomalie stają się wykrywalne poprzez określenie odchylenia od modelu.

Ostatnio zauważalny jest wzrost zainteresowania opracowywaniem podejść uczenia głębokiego dla detekcji wartości odstających. Motywacją jest brak skutecznych metod dla zadań detekcji wartości odstających, które obejmują złożone dane. Podobnie jak w innych przypadkach zastosowania uczenia głębokiego, celem głębokiej detekcji wartości odstających jest złagodzenie obciążenia manualnego dobierania cech i umożliwienie budowy efektywnych, skalowalnych rozwiązań. Jednak w odróżnieniu od nadzorowanego uczenia głębokiego, jest mniej jasne, jakie są użyteczne cele uczenia reprezentacji dla głębokiej detekcji wartości odstających, ze względu na głównie nienadzorowany charakter problemu.

Główne podejścia do głębokiej detekcji wartości odstających obejmują warianty głębokich autoenkoderów [234], głęboką klasyfikację jednoklasową [235, 236], metody oparte na modelach generatywnych, GANy [237]. W porównaniu do tradycyjnych metod detekcji anomalii, gdzie reprezentacja cech jest ustalana a priori (np. poprzez jądro funkcji cech), te podejścia mają na celu naukę mapy cech danych $\phi_\omega : x \to \phi_\omega(x)$, gdzie ϕ_ω jest głęboką siecią neuronową z parametrami wagowymi w, jako część ich celu uczenia się.

Obecnie detekcja wartości odstających ma liczne zastosowania w różnych dziedzinach, takich jak cyberbezpieczeństwo [238, 239], fizyka [240, 241] diagnozy medyczne [242, 243], ekonomia [244] utrzymanie infrastruktury [245], wykrywanie nowości akustycznych [234] i wiele innych.

Wykrywanie anomalii to także kluczowa technika stosowana w dziedzinie wykrywania usterek maszyn. Polega na monitorowaniu danych z maszyn lub systemów, aby identyfikować nietypowe lub nieprawidłowe wzorce, które mogą wskazywać na potencjalne problemy. Podejście to stosowane jest w wykrywaniu usterek i uszkodzeń maszyn, takich jak silniki, urządzenia mechaniczne czy systemy produkcyjne [219, 246]. Główną zaletą przemawiającą za stosowaniem analizy anomalii do wykrywania usterek maszyn jest możliwość pracy na nieoznaczonych danych, czyli danych pobieranych bezpośrednio z czujników maszyny i niepoddawanych dodatkowemu przetwarzaniu. Oznacza to, że możliwe jest interpretowanie i wykrywanie usterek na podstawie różnic pomiędzy danymi podczas prawidłowej pracy maszyn i urządzeń oraz przy awariach. Jest to duża zaleta, ponieważ sztuczne wywoływanie awarii jest często bardzo kosztowne lub wręcz niemożliwe, ponadto bardzo często nie jest możliwe przewidzenie wszystkich potencjalnych typów awarii jakie mogą wystąpić.

Dane pomiarowe dostępne w wymienionych dziedzinach stale rosną pod względem wielkości. Podejścia ewoluują w kierunku obsługi złożonych typów danych, takich jak obrazy, filmy, dźwięki, tekst, wielowymiarowe szeregi czasowe czy sekwencje biologiczne.

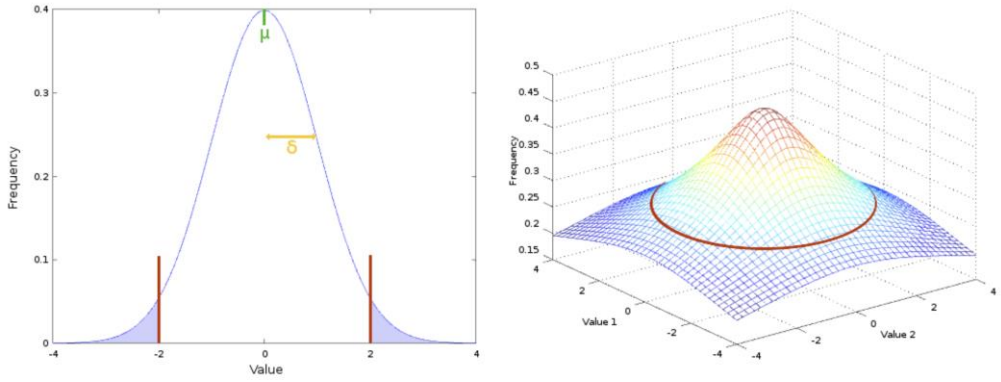
3.5.1. Definicja anomalii

Pod pojęciem anomalii rozumiemy obserwację, która znacząco odbiega od pewnego pojęcia normalności.

Niech X ⊆ R^D będzie przestrzenią danych zadaną przez pewne zadanie lub zastosowanie. Pojęcie normalności definiujemy jako rozkład P^+ na X, który jest normalnym zachowaniem w danym zastosowaniu lub zadaniu. Obserwacja znacznie odbiegająca od takiego stanu normalności, czyli anomalia, to punkt lub zbiór punktów $x \in X$ danych, który leży w obszarze niskiego prawdopodobieństwa według P^+. Zakładając, że P^+ ma odpowiadającą funkcję gęstości prawdopodobieństwa $p^+(x)$, możemy zdefiniować zbiór anomalii jako [324]:

$$A = (x \in X | p^+(x) \leq \tau), \quad \tau \geq 0 \tag{1}$$

gdzie τ to pewna granica, taka, że prawdopodobieństwo A według P^+ jest wystarczająco małe.



Rysunek 20 Zobrazowanie marginesu anomalii (https://www.rudikershaw.com/articles/anomaly-detection, dostęp 03.02.2024)

Na rysunku 20 po lewej znajduje się krzywa dzwonowa z jedną zmienną. Dwie czerwone linie reprezentują przykładową granicę anomalii. Wszystko, co znajduje się wewnątrz granic (dane o dużej częstotliwości występowania), uważane jest za normalne, a wszystko na zewnątrz (w jasnoniebieskim obszarze), to anomalia. Położenie tego progu jest arbitralne. W wypadku dodania drugiej zmiennej, otrzymujemy wykres 3D, jak na rysunku 33 po prawej. Jeśli mamy więcej niż dwie zmienne, granica anomalii staje się trudna to zwizualizować, ale koncepcja jest taka sama w przestrzeni wielowymiarowej.

3.5.2. Metody detekcji anomalii

W literaturze znaleźć można wiele badań proponujących różne podejścia do wykrywania anomalii. Podejścia te można podzielić na modele rekonstrukcyjne, metody klasyfikacji

jednoklasowej oraz metody estymacji gęstości i modele probabilistyczne. W następnej sekcji przybliżę podstawowe założenia podejść opartych o techniki uczenia maszynowego.

Modele rekonstrukcyjne

Modele rekonstrukcyjne należą do najwcześniejszych i najczęstszych podejść sieci neuronowych do detekcji anomalii [247, 248]. Metody oparte na rekonstrukcji przygotowują model, który jest zoptymalizowany do dobrego reprodukowania normalnych zestawów danych. Model wykrywa anomalie, ponieważ nie są one dokładnie odtwarzane w ramach nauczonego modelu. Większość tych metod ma czysto geometryczną motywację (na przykład PCA lub deterministyczne AE), ale niektóre warianty probabilistyczne ujawniają związek z estymacją gęstości prawdopodobieństwa.

Założenia metod rekonstrukcyjnych prezentują się następująco [247]. Niech $:X \to X, x \to \phi_\theta(x)$ będzie odwzorowaniem cech przestrzeni danych X na siebie, które składa się z funkcji kodującej $\phi_e: X \to Z$ (enkoder) i funkcji dekodującej $\phi_d: Z \to X$ (dekoder). Dalej $\phi_e: X \to Z$, gdzie $\phi_\theta \equiv (\phi_e \circ \phi_d)$, gdzie θ obejmuje parametry zarówno enkodera, jak i dekodera. Przestrzeń Z nazywamy przestrzenią ukrytą, a $\phi_e(x) = z$ reprezentacją ukrytą x.

Celem rekonstrukcji w kontekście uczenia maszynowego jest znalezienie funkcji ϕ_θ, takiej, że:

$$\phi_{\theta(x)} = \phi_d\big(\phi_e(x)\big) = \hat{x} \approx x \tag{2}$$

Oznacza to znalezienie transformacji kodującej i dekodującej, dzięki którym x zostanie zrekonstruowane z minimalnym błędem, zazwyczaj mierzonym w odległości euklidesowej. Dla danych nieetykietowanych $x_1, \dots, x_n \in X$ cel funkcji rekonstrukcji jest określony jako:

$$\min_\theta \frac{1}{n} \sum_{i=1}^{n} \|x_i - (\phi_d \circ \phi_e)_\theta(x_i)\|^2 + R \tag{3}$$

gdzie R oznacza różne formy regularyzacji wprowadzane przez różne metody, na przykład dotyczące parametrów θ, struktury transformacji kodującej i dekodującej, lub geometrii przestrzeni ukrytej Z.

Model trenowany w oparciu o cel rekonstrukcji musi wyodrębniać istotne cechy i charakterystyczne wzorce z danych w swoim kodowaniu tak, aby dekodowanie z kompresowanej reprezentacji ukrytej osiągnęło niski błąd rekonstrukcji. Zakładając, że dane treningowe $x_1, \dots, x_n \in X$ zawierają głównie punkty normalne, możemy oczekiwać, że model oparty na rekonstrukcji będzie generował niski błąd rekonstrukcji dla instancji normalnych i wysoki błąd rekonstrukcji dla anomalii. Z tego powodu wystąpienie anomalii jest zazwyczaj bezpośrednio określane przez błąd rekonstrukcji:

$$s(x) = \|x - (\phi_d \circ \phi_e)_\theta(x)\|^2 \tag{4}$$

Dla modeli, które nauczyły się pewnej prawdziwej struktury lub prototypowej reprezentacji, wysoki błąd rekonstrukcji wykryje instancje anomalii. Większość metod rekonstrukcji nie podąża za żadnym motywem probabilistycznym, a punkt x zostaje oznaczony jako nietypowy po prostu dlatego, że nie odpowiada swojej „idealizowanej" reprezentacji $\phi_d(\phi_e(x)) = \hat{x}$ uzyskanej w procesach kodowania i dekodowania. Jednakże niektóre metody rekonstrukcji takie jak PCA [230] czy autoenkodery wariacyjne VAE [249], mają również interpretacje probabilistyczne. Te metody są związane z estymacją gęstości (przy założeniu określonych struktur ukrytych), zazwyczaj w sensie, że wysoki błąd rekonstrukcji wskazuje na obszarach o niskiej gęstości i vice versa.

Autoenkodery (AE)

Autoenkodery jak opisywaliśmy w rozdziale 0 to modele rekonstrukcji, które wykorzystują sieci neuronowe do kodowania i dekodowania danych. Były one badane już na wczesnym etapie swojej historii pod kątem detekcji anomalii. Obecnie głębokie autoenkodery są jednymi z najczęściej stosowanych metod w obszarze głębokiej detekcji anomalii [220, 234 ,243, 246] prawdopodobnie ze względu na swoją długą historię i łatwe do użycia standardowe warianty.

Powszechnym sposobem regularyzacji autoenkoderów jest odwzorowywanie do niższego wymiaru $\phi_e(x) = z \in Z$ poprzez enkoder (tzw. wąskie gardło), co narzuca kompresję danych i efektywnie ogranicza wymiarowość lub podprzestrzeń do nauczenia. Oprócz „wąskiego gardła" w literaturze wprowadzono wiele różnych sposobów na regularyzację autoenkoderów. Autoenkodery odszumiające (DAE) [250, 251] eksponują dane wejściowe zakłócone szumem $\tilde{x} = x + \varepsilon$, które następnie są trenowane do odtwarzania

oryginalnych danych. DAE umożliwiają więc określenie modelu szumu dla ε. Podążając za koncepcjami kodowania rzadkiego, autoenkodery rzadkie [168, 176, 252] regularyzują kod ukryty w kierunku rzadkości, na przykład poprzez penalizację L1 Lasso [253]. W sytuacjach, w których dane szkoleniowe są już zanieczyszczone szumem lub nieznanymi anomaliami, zaproponowano odporną głęboką wersję autoenkoderów [254], która dzieli dane na dobrze reprezentowane i zanieczyszczone części. Autoenkodery CAE [255] proponują ukaranie funkcji aktywacji enkodera normą Frobeniusa względem danych wejściowych, aby uzyskać gładką i bardziej odporną reprezentację ukrytą. Takie metody regularyzacji wpływają na geometrię i kształt podprzestrzeni lub rozmaitości, które są uczone przez AE, na przykład, poprzez narzucanie pewnego stopnia gładkości lub wprowadzanie odporności wobec pewnych typów zniekształceń czy zakłóceń wejściowych. Dlatego te wybory regularyzacji powinny ponownie odzwierciedlać konkretne założenia danego zadania wykrywania anomalii.

Wreszcie, inne warianty autoenkoderów, które zostały zastosowane do detekcji anomalii, obejmują autoenkodery konwolucyjne [174, 251], autoenkodery oparte na RNN [171, 220] czy autoenkodery zespołowe [256]. Autoenkodery zostały również wykorzystane w podejściach dwuetapowych, które używają autoenkoderów do redukcji wymiarowości i stosują tradycyjne metody do klasyfikacji [220, 234, 258].

Sieci GAN

Generatywne sieci przeciwników GAN (ang. Generative adversarial network) definiują problem nauki docelowego rozkładu cech jako zagadnienie gry o sumie zerowej [237]. W tym podejściu model generatywny jest szkolony w konkurencji z przeciwnikiem, który pobudza go do generowania próbek, których rozkład jest podobny do rozkładu treningowego. Model GAN składa się z dwóch sieci neuronowych: sieci generatora $\phi_\omega: Z \to X$ i sieci dyskryminatora $\psi_{\omega'}: X \to (0,1)$. Obie sieci rywalizują ze sobą, aby dyskryminator był szkolony do rozróżniania $\phi_\omega(z)$ od $x \sim P^+$, gdzie $z \sim Q$. Generator jest szkolony do oszukiwania sieci dyskryminatora, co powoduje generowanie próbek bardziej podobnych do docelowego rozkładu. Osiągnięcie tego celu adwersarialnego można zapisać jako:

$$\min_{\omega} \min_{\omega'} E_{x \sim P^+}[\log \psi_{\omega'}(x)] + E_{z \sim Q}[\log(1 - \psi_{\omega'}(\phi_\omega(z))] \tag{5}$$

Szkolenie jest zazwyczaj przeprowadzane za pomocą przemiennego schematu optymalizacji. Istnieje wiele wariantów GAN, np. Wasserstein GAN [259, 260], które są często

używane w metodach detekcji anomalii. Z powodu swojej konstrukcji modele GAN nie oferują sposobu przypisania punktom w przestrzeni wejściowej prawdopodobieństwa. Inne podejścia stosują optymalizację w celu znalezienia punktu $\tilde{z}$ w przestrzeni ukrytej $\tilde{x}$, takiego, że $\tilde{x} \approx \phi_\omega$: $(\tilde{z})$ dla punktu testowego x. Schlengl et al. [261] zaproponowali model AnoGAN gdzie zastosowali warstwę pośrednią dyskryminatora, $f_{\omega'}$, i użyli kombinacji straty rekonstrukcji $\|\tilde{x} - \phi_\omega(\tilde{z})\|$ i straty dyskryminacji $\|f_{\omega'}(\tilde{x}) - f_{\omega'}(\phi_\omega(\tilde{z}))\|$ jako wyniku anomalii. GANomaly [262] używał większej metryki odległości od nauczonego rozkładu danych w ograniczonej przestrzeni ukrytej do wnioskowania jako wskaźnik nietypowości danej obrazu. Efficient-GAN [263] zdefiniował funkcję oceny, aby sprawdzić, czy odtworzone dane posiadają podobne cechy w dyskryminatorze jak próbka rzeczywista. ALAD [264] wykorzystywał błędy rekonstrukcji między parą próbek rzeczywistych i ich rekonstrukcją, aby określić, czy próbka danych jest nietypowa. f-AnoGAN zaproponowany przez [265] wykrył próbki nietypowości i (poprzez dodanie reszty do cech dyskryminatora) dostarczył niezawodnej etykiety dla anomalii, jak również wysokie oceny anomalii dla nietypowych obrazów i niskie oceny dla zdrowych obrazów.

Autoenkodery VAE

Oprócz opisywanych w rozdziale 0 deterministycznych wariantów AE, zaproponowano również autoenkodery probabilistyczne, które nawiązują do estymacji gęstości. Najbardziej zbadaną klasą autoenkoderów probabilistycznych są VAE (ang. Variational Autoencoders) [249].

VAE uczą głębokie modele zmiennych ukrytych, gdzie wejścia *x* są parametryzowane na próbkach zmiennych ukrytych $z \sim Q$ za pomocą pewnej sieci neuronowej, aby nauczyć ją rozkładu $p_\theta(x|z)$ takiego, że $p_\theta(x) \sim p^+(x)$. Powszechnie przyjmuje się, że Q jest izotropowym wielowymiarowy rozkładem Gaussa, a sieć neuronowa $\phi_{d,\omega} = (\mu_\omega, \sigma_\theta)$ (dekoder) o wagach ω parametryzuje średnią i wariancję rozkładu Gaussa, czyli

$$p_\theta(x|z) \sim N(x, \mu_\omega(z), \sigma_\omega^2(z)I) \tag{6}$$

Sieć enkodera $\phi_{e,\omega'}$ wprowadza się w celu parametryzacji rozkładu wariacyjnego $p_{\theta'}(z|x)$, z θ' zawartym w wynikach $\phi_{e,\omega'}$, w celu przybliżenia ukrytego pierwotnego $p(z|x)$. Cały model jest następnie optymalizowany metodą wariacyjną Bayesa:

$$\max_{\theta,\theta'} -D_{KL}(q_{\theta'}(z|x)||p(z)) + E_{q_{\theta'}(z|x)}[\log p_\theta(x|z)] \quad (7)$$

Optymalizacja przebiega za pomocą stochastycznego spadku gradientowego wariacyjnego [249]. Po wytrenowaniu VAE można oszacować $p_\theta(x)$ za pomocą próbkowania Monte Carlo z rozkładu $p(z)$ i $E_{z\sim p(z)}[p_\theta(x|z)]$. Użycie wyniku oceny bezpośrednio do analizy anomalii ma dogodne teoretyczne uzasadnienie, ale eksperymenty pokazały, że zazwyczaj działa gorzej niż użycie prawdopodobieństwa rekonstrukcji [266], które warunkuje na *x*, aby oszacować $E_{q_{\theta'}(z|x)}[\log p_\theta(x|z)]$.

Różne formy VAE zostały zaimplementowane w różnych formach do wykrywania anomalii. Yao et al. [267], zaproponowali algorytm oparty o VAE, który stosowali do wydobycia wartościowych cech do zadań nienadzorowanego wykrywania anomalii. Podobne podejścia stosowali w swoich pracach Xu et al. [268] i Nalisnick et al. [269].

Inną klasą probabilistycznych autoenkoderów, które zostały zastosowane do detekcji anomalii, są autoenkodery adversarialne (AAE) [234]. Przy użyciu straty adwersarialnej do regularyzacji i dopasowywania rozkładu kodowania ukrytego, AAE mogą używać dowolnego arbitralnego *p(z)*, o ile próbkowanie jest możliwe.

Stosowane są także modele hybrydowe łączące VAE z innymi typami sieci neuronowych, na przykład w danych sekwencyjnych wielomodalnych z użyciem sieci LSTM [270, 271], czy połączenie VAE-SVDD [272, 273].

Klastrowanie prototypowe

Metody klastrowania, stanowią kolejne podejście do detekcji anomalii opartej na rekonstrukcji. Błąd rekonstrukcji jest tutaj zazwyczaj określany jako odległość punktu od najbliższego prototypu, który idealnie został nauczony reprezentować normalny rozkład danych. Metody klastrowania prototypowego [274] obejmują dobrze znane algorytmy k-median, k-medoidów, VQ, gdzie zazwyczaj są stosowane w przestrzeni wejściowej X. Metody k-średnich były z powodzeniem stosowane w wykrywaniu anomalii [275, 276]. Badano także warianty z jądrem k-średnich [277] i stosowano to podejście w detekcji anomalii [278]. Modele Mieszaniny Gaussa (GMM) z ograniczoną liczbą k składników zostały użyte do klastrowania prototypowego anomalii [279]. Tutaj odległość od każdego klastra

(lub składnika danych) jest określana przez odległość Mahalanobisa, która jest definiowana przez macierz kowariancji odpowiedniego składnika mieszaniny gaussowskiej. Innym z podejściem do wykrywania anomalii są tzw. lasy izolacyjne (ang. Isolation forest) po raz pierwszy zaproponowane przez Fei et al. [280]. Algorytm ten stosuje drzewa binarne i opiera się na cechach anomalii, tj. ich nieliczności i odmienności, aby je wykrywać.

Niedawno wprowadzono także podejścia do klastrowania oparte na głębokim uczeniu [281, 282]. Podejście to z sukcesem zastosowane zostało w detekcji anomalii [257, 258, 283, 284]. Tak jak w wielu zagadnieniach korzystających z głębokich technik uczenia problemem w głębokim klastrowaniu jest efektywna regularyzacja w celu przeciwdziałania zanikowi mapy cech [285]. Należy zwrócić uwagę na fakt, że w przypadku głębokich autoenkoderów błąd rekonstrukcji jest mierzony w przestrzeni wejściowej X po dekodowaniu, natomiast w metodach głębokiego klastrowania, błąd rekonstrukcji mierzony jest w przestrzeni ukrytej Z. Zatem zanik cech ukrytych, czyli stały enkoder $\phi_e \equiv c \in Z$ skutkowałby stałym dekodowaniem (średnią danych przy optymalnych warunkach) dla autoenkodera, co generalnie jest suboptymalnym rozwiązaniem. Z tego powodu autoenkodery wydają się mniej podatne na zanik cech, chociaż również zaobserwowano, że zbiegają się do złych optimum lokalnych podczas optymalizacji SGD, zwłaszcza jeśli używają składników obciążenia.

Analiza składowych głównych

Analiza składowych głównych PCA (ang. Principal component analysis) [229] to technika liniowej redukcji wymiarowości stosowana w eksploracyjnej analizie danych, wizualizacji oraz przetwarzaniu wstępnym danych. Jednym z często spotykanych sposobów formułowania celu PCA jest poszukiwanie ortogonalnej bazy W w przestrzeni danych $X \subseteq R^D$, która maksymalizuje empiryczną wariancję (scentralizowaną) danych $x_1, \dots, x_n \in X$

$$\min_W \sum_{i=1}^{n} \|x_i - W^T W x_i\|^2, \quad takie\ że\ WW^T = I \tag{8}$$

Rozwiązanie tej funkcji celu prowadzi do znanego problemu własnościowego [286], ponieważ optymalna baza jest dana przez wektory własne macierzy kowariancji empirycznej, gdzie odpowiednie wartości własne odpowiadają wariancjom składowych. Komponenty składowe $d \leq D$, które wyjaśniają większość wariancji (tzw. główne składowe) są następnie

określane przez *d* wektorów własnych, które mają największe wartości własne. Z perspektywy rekonstrukcji, cel polega na znalezieniu ortogonalnej projekcji $W^T W$ do d-wymiarowej podprzestrzeni liniowej, tak aby błąd rekonstrukcji był minimalizowany wg wzoru:

$$\min_{W} \sum_{i=1}^{n} \|x_i - W^T W x_i\|^2 , \quad takie\ że\ WW^T = I \tag{9}$$

Zatem PCA optymalnie rozwiązuje cel rekonstrukcji dla liniowego kodera $\phi_e(x) = Wx = z$ oraz transponowanego liniowego dekodera $\phi_d(z) = W^T$ z ograniczeniem $WW^T = I$. Dla liniowego PCA, można zidentyfikować jego interpretację probabilistyczną [287]. Rozkład danych wynika z liniowej transformacji $X = W^T Z + \varepsilon$ d-wymiarowego rozkładu Gaussa $Z \sim N(0,1)$, tak że $P \equiv N(0, W^T W + \sigma^2 I)$. Maksymalizowanie prawdopodobieństwa tego rozkładu Gaussowskiego względem parametrów kodowania i dekodowania W ponownie prowadzi do PCA jako optymalnego rozwiązania [287]. Dlatego też, PCA zakłada, że dane znajdują się na d-wymiarowej elipsoidzie osadzonej w przestrzeni danych $X \subseteq R^D$. Standardowe PCA zapewnia zatem ilustracyjny przykład związków między estymacją gęstości, a rekonstrukcją. Liniowe PCA jest oczywiście ograniczone do kodowań danych, które wykazują tylko liniowe korelacje cech. kPCA [230] wprowadziło nieliniową generalizację analizy składowych poprzez rozszerzenie celu PCA na nieliniowe odwzorowania cech jądra. Szereg prac badawczych [230, 288-290] zastosowało PCA do wykrywania anomalii.

Klasyfikatory jednoklasowe

Klasyfikacja jednoklasowa [285, 291], opiera się na podejściu dyskryminacyjnym do detekcji anomalii. Metody oparte na klasyfikacji jednoklasowej starają się unikać pełnej estymacji gęstości. Zamiast tego te metody mają na celu bezpośrednie wytworzenie granicy decyzyjnej, która prowadzi do niskiego błędu, gdy jest stosowana do niewidzianych danych lub inaczej nauczenie się granicy decyzyjnej, która odpowiada poziomowi gęstości pożądanego rozkładu danych normalnych P^+.

Możemy traktować klasyfikację jednoklasową jako szczególnie trudny problem klasyfikacji, a mianowicie jako klasyfikację binarną, w której mamy tylko (lub prawie tylko) dostęp do danych z jednej klasy – klasy normalnej. W tym nierównoważonym ustawieniu cel klasyfikacji jednoklasowej polega na nauczeniu się granicy decyzyjnej jednoklasowej, która minimalizuje:

- Pominięte lub nieodkryte prawdziwe anomalie (tj. współczynnik pominięcia lub błąd II rodzaju).

- fałszywe alarmy dla prawdziwych instancji normalnych (tj. współczynnik fałszywego alarmu lub błąd I rodzaju)

Osiągnięcie niskiego współczynnika fałszywego alarmu jest koncepcyjnie proste: przy dostatecznej liczbie punktów danych normalnych można by po prostu narysować pewną granicę obejmującą wszystkie istotne dane. Problemem jest jednak określenie tej granicy dostatecznie blisko danych bez anomalii. Z tego powodu zazwyczaj a priori określa się jakiś docelowy współczynnik fałszywego alarmu $\alpha \in [0,1]$, dla którego następnie dąży się do minimalizacji współczynnika pominięcia. Klasyfikacja jednoklasowa polega więc na minimalizowaniu wskaźnika błędu pominięcia dla określonej wartości wskaźnika fałszywego alarmu przy braku dostępu do anomalii.

Możemy wyrazić powyższe uzasadnienie w terminach ryzyka klasyfikacji binarnej [292]. Niech $Y \in \{\pm 1\}$ będzie zmienną losową z klasy, gdzie $Y = +1$ oznacza dane normalne, a $Y = -1$ oznacza punkty nietypowe. Dzięki czemu możemy zidentyfikować rozkład danych normalnych jako $P^+ \equiv P_{X|Y=+1}$, a rozkład anomalii jako $P^- \equiv P_{X|Y=-1}$. Ponadto, niech $\ell: R \times \{\pm 1\} \to R$ będzie funkcją straty klasyfikacji binarnej, a $f: X \to R$ niech będzie pewną rzeczywistą funkcją wyników. Ryzyko klasyfikacji f w kontekście straty wynosi wtedy:

$$R(f) = E_{X \sim P^+}[\ell(f(X), +1)] + E_{X \sim P^-}[\ell(f(X), -1)] \tag{10}$$

Minimalizowanie drugiego składnika – oczekiwanej straty przy klasyfikowaniu prawdziwych anomalii jako normalnych – odpowiada minimalizacji współczynnika pominięcia. Dla pewnych danych nieoznaczonych $x_1, \dots, x_n \in X$ i ewentualnie dodatkowych oznaczonych danych $(\widetilde{x_1}, \widetilde{y_1}, \dots, \widetilde{x_m}, \widetilde{y_m})$, możemy zastosować zasadę minimalizacji empirycznej straty, aby uzyskać:

$$\min_f \frac{1}{n} \sum_{i=1}^{n} \ell(f(x_i), +1) + \frac{1}{m} \sum_{j=1}^{m} \ell\left(f\left(\widetilde{x_j}\right), \widetilde{y_j}\right) + R \tag{11}$$

Powyższe równanie określa cel klasyfikacji jednoklasowej w sposób empiryczny. Drugi składnik równania to pusta suma w trybie nienadzorowanym. Dodajemy R jako dodatkowy

składnik, aby oznaczyć i uwzględnić regularyzację, która może przyjąć różne formy, zależnie od założeń dotyczących f, ale także od P^-. Ogólnie rzecz biorąc, regularyzacja $R = R(f)$ ma na celu zminimalizowanie współczynnika pominięcia (np. poprzez minimalizację objętości i założenia dotyczące P^-) oraz poprawę uogólniania (np. przez wygładzanie f). Ponadto, zauważmy, że $y = +1$ w pierwszym składniku uwzględnia założenie, że n-nieoznaczonych punktów danych treningowych są punktami normalnymi. To założenie można jednak dostosować poprzez odpowiednio wybrane funkcje straty i regularyzacji, na przykład wymagając, aby pewien odsetek danych nieoznaczonych został źle sklasyfikowany. Pozwala to uwzględnić założenie dotyczące współczynnika zanieczyszczenia η lub osiągnięcia pewnego docelowego współczynnika fałszywego alarmu α.

Jednoklasowa klasyfikacja oparta na jądrze

Jednymi z najbardziej znanych metod klasyfikacji jednoklasowej opartej na jądrze są OC-SVM [231] i SVDD [232]. Niech $;: X \times X \to R$ będzie pewnym dodatnio określonym jądrem o skojarzonej przestrzeni Hilberta z jądrem reprodukującym F_k i odpowiadającej mu mapie cech $\phi_k: X \to F_k$, więc $k(x, \widetilde{x}) = \langle \phi_k(x), \phi_k(\widetilde{x}) \rangle$ dla wszystkich $x, \tilde{x} \in X$. Celem metody SVDD jest znalezienie otaczającej dane hipersfery o minimalnej objętości. Pierwotny problem SVDD to:

$$\min_{R,C,\xi} R^2 + \frac{1}{vn} \sum_{i=1}^{n} \xi_i \tag{12}$$

Takie że: $\|x_i - c\|^2 \leq R^2 + \xi_i, \quad \xi_i \geq 0 \quad \forall i$

SVDD używa modelu hipersfery $f_\theta(x) = \|\phi_k(x) - c\|^2 - R^2$ zdefiniowanego w przestrzeni cech F_k. W porównaniu do tego, obiektyw OC-SVM polega na znalezieniu hiperpłaszczyzny $\omega \in F_k$, która oddziela dane w przestrzeni cech F_k z maksymalnym marginesem:

$$\min_{\omega,\rho,\xi} \frac{1}{2} \|\omega\|^2 - \rho + \frac{1}{vn} \sum_{i=1}^{n} \xi_i \tag{13}$$

Takie że: $\rho - \langle \phi_k(x_i), \omega \rangle \leq \xi_i, \quad \xi_i \geq 0 \quad \forall i$

Zatem OC-SVM używa modelu liniowego $f_\theta = \rho - \langle \phi_k(x_i), \omega \rangle$ w przestrzeni cech F_k z parametrami modelu $\theta = (\omega, \rho)$. Margines do danych prawdziwych wynosi $(\frac{\rho}{\|\omega\|})$ i jest maksymalizowany poprzez maksymalizację ρ, gdzie $\|\omega\|$ działa jako normalizator.

Zarówno SVDD jak i OC-SVM można rozwiązać w ich odpowiednich formułach dualnych, które są programami kwadratowymi i obejmują tylko iloczyny skalarne (mapa cech ϕ_k jest ukryta). Dla standardowego jądra gaussowskiego (lub dowolnego jądra o stałej normie $k(x, x) = c > 0$), SVDD i OC-SVM są równoważne [291].

Przestrzenie cech tworzone przez jądra znacznie zwiększają możliwości opisowe metod jednoklasowych i pozwalają na uczenie dobrze działających modeli działających na złożonych danych. Badacze przetestowali wiele wariantów jednoklasowych klasyfikatorów opartych na jądrze, takich jak wielosferowe SVDD [278], hierarchiczne osadzanie estymacji poziomu gęstości [293], wielojądrowe uczenie dla OC-SVM [294], opis danych rozproszonych [295] i bayesowskich [296], inkrementalne uczenie [297], lub odporne [298] warianty.

Klasyfikacja jednoklasowa oparta na głębokim uczeniu

Wybór typu jądra i ręczne separowanie istotnych cech, może być trudne i szybko stać się niepraktyczne dla skomplikowanych danych. Metody klasyfikacji jednoklasowej oparte na głębokim uczeniu starają się pokonać te wyzwania, ucząc użytecznych map cech sieci neuronowych $\phi_\omega: X \to Z$ z danych lub przenosząc takie sieci z zadań pokrewnych. Głębokie SVDD [285, 299, 300] oraz warianty DL OC-SVM [236, 301] korzystają z modelu hipersfer $f_\theta(x) = \|\phi_\omega(x) - c\|^2 - R^2$ i modelu liniowego $f_\theta(x) = \rho - \langle \phi_\omega(x), \omega \rangle$ z jawnie określonymi mapami cech. Te metody są zazwyczaj optymalizowane za pomocą wariantów SGD [301], co, w połączeniu z równoległością GPU, pozwala na skalowanie do dużych zbiorów danych.

Jednoklasowy Deep SVDD [285] został wprowadzony jako prostszy wariant w porównaniu z użyciem modelu neuronowego hipersfery, co stawia sobie za cel:

$$\min_{\omega, c} \frac{1}{n} \sum_{i=1}^{n} \|\phi_\omega(x_i) - c\|^2 + R \tag{14}$$

Tu transformacja sieci neuronowej $\phi_\omega(\cdot)$ jest uczona w celu zminimalizowania średniego kwadratu odległości wszystkich punktów danych do centrum $c \in Z$. Optymalizacja tego uproszczonego celu okazała się szybka i skuteczna w wielu sytuacjach [285]. Jednoklasowe głębokie SVDD można również interpretować jako jednoprzykładową głęboką metodę klastrowania.

Badacze sprawdzili także możliwość zastosowania innych metod głębokiego uczenia w zagadnieniu nienadzorowanego wykrywania anomalii. Proponowane modele zwykle uczone są do rekonstrukcji normalnych reprezentacji złożonych danych. Wykrywanie anomalii odbywa się na podstawie błędu rekonstrukcji danych nie biorących udziału w uczeniu. Sieci RNN [303], LSTM [271] oraz opisywane wcześniej sieci Autoenkoderów (rozdział 0), GAN (rozdział 0) oraz ich wariacje [219, 254, 264, 270, 272, 273] są szeroko stosowane do danych sekwencyjnych, takich jak tekst, dźwięk i szeregi czasowe. CNN odgrywa główną rolę w przypadku danych nieciągłych, takich jak obrazy, sieci i sensory [236, 251, 304].

Ważnym zagadnieniem w klasyfikacji jednoklasowej opartej na głębokim uczeniu jest to, jak sensownie regulować mapę cech $\phi_\omega \equiv c$. Bez regularyzacji, zagadnienia minimalizacji objętości lub maksymalizacji obszaru marginesu można by łatwo rozwiązać za pomocą stałego odwzorowania [285]. Możliwe rozwiązania to użycie prawdziwych [305], pomocniczych [306] lub sztucznych [306] negatywnych przykładów w szkoleniu, dodanie terminu rekonstrukcji lub ograniczeń architektonicznych [285], karanie wariancji osadzenia i użycie zaszumionych danych [307], Pseudolabeling [307, 308], zamrożenie osadzenia [235, 236] lub integracja pewnego założenia o rozmaitości [309].

Metody głębokiej klasyfikacji jednoklasowej zazwyczaj oferują większą elastyczność modelowania i umożliwiają naukę lub transfer cech istotnych dla skomplikowanych danych. Zazwyczaj jednak wymagają więcej danych, aby być skutecznymi lub muszą polegać na pewnym wcześniejszym modelu wzorcowym (na przykład na jakiejś wstępnie nauczonej sieci). Niemniej jednak zasada leżąca u podstaw metod klasyfikacji jednoklasowej — próba określenia dyskryminacyjnej granicy jednoklasowej podczas nauki — pozostaje niezmieniona, niezależnie od tego, czy używana jest płytka lub głęboka mapa cech.

3.5.3. Podsumowanie przeglądu metod detekcji anomalii

Wykrywanie anomalii jest obecnie obiektem szerokiego zainteresowania teoretycznego i praktycznego badaczy. Rozwiązania automatycznego wykrywania wartości odstających zostało zastosowane z powodzeniem w wielu dziedzinach techniki. Jak wykazał przegląd literatury dziedzina ta rozwija się w oparciu o kilka głównych kierunków: klasyczne podejścia oparte o analizę skupień i metody oparte na jądrach oraz obecnie silnie rozwijane podejście uczenia głębokiego reprezentacji danych.

Jak wykazała analiza literaturowa, pomimo powstawania bardzo licznych publikacji w temacie analizy anomalii, metodologia ta nie była do tej pory stosowana w analizie danych pomiarowych procesu FHD. Wykrywanie anomalii może posłużyć jako dodatkowe narzędzie w kontroli procesu drążenia otworów wentylacyjnych. Nienadzorowany charakter metod wykrywania anomalii ma potencjał do analizy danych bez konieczności symulowania dużej ilości usterek i ma możliwość wykrycia sytuacji wcześniej nie zaobserwowanych. Wyniki tego typu analiz posłużyć mogą do identyfikowania błędnie wykonywanych otworów, co może ograniczyć ilość czynności kontrolnych po wierceniu. Na podstawie zidentyfikowanych anomalii występujących podczas drążenia, w późniejszych badaniach można będzie podejmować próby korelacji typów anomalii z wadami detali, w celu łatwiejszego identyfikowania przyczyn ich występowania. Anomalie drążeń mogą wynikać także ze zużywania się lub uszkodzenia części elektrodrążarki. Wyniki z przetwarzania danych mogą posłużyć za wsad do identyfikacji trendów zmienności parametrów obróbki, co może być jednym z typów danych wejściowych dla tworzonych w przyszłości systemów predykcyjnego utrzymania ruchu maszyn.

4. Badania eksperymentalne procesu FHD

W części eksperymentalnej pracy wykonywano drążenia na elementach odwzorowujących prawdziwe częściach silników odrzutowych. Zastosowano materiały, budowę próbek testowych oraz zestawy parametrów procesu, które pozwoliły na opracowanie i potwierdzenie skuteczności tworzonych metod.

Rysunek 21 Maszyna FHD MKG 1000 JR Automation

Do drążenia elektroerozyjnego wykorzystano drążarkę sterowaną numerycznie MKG 1000 wyprodukowaną przez firmę JR Automation (Rysunek 21). Urządzenie pozwala na precyzyjne operowanie w 5ciu osiach, co umożliwia wykonywanie skomplikowanych kształtów w trakcie wykonywania całych wierszy otworów chłodzących oraz charakteryzuje się relatywnie niedużymi wymiarami w porównaniu z rynkowo dostępnymi konstrukcjami do FHD.

Maszyna pozwala na ustawianie parametrów pracy. Napięcie i natężenie zmieniać można w przedziale: 0-100V oraz 0-50A. Czas pracy (T_{on}) oraz czas płukania (T_{off}) regulowany może być w przedziale 0-999µs z rozdzielczością do pojedynczych mikrosekund. Prędkość obrotowa wrzeciona może być regulowane w zakresie 0-500 obr/min. Dielektryk w postaci wody dejonizowanej o stałej konduktywności na poziomie 10mS/cm jest dostarczany

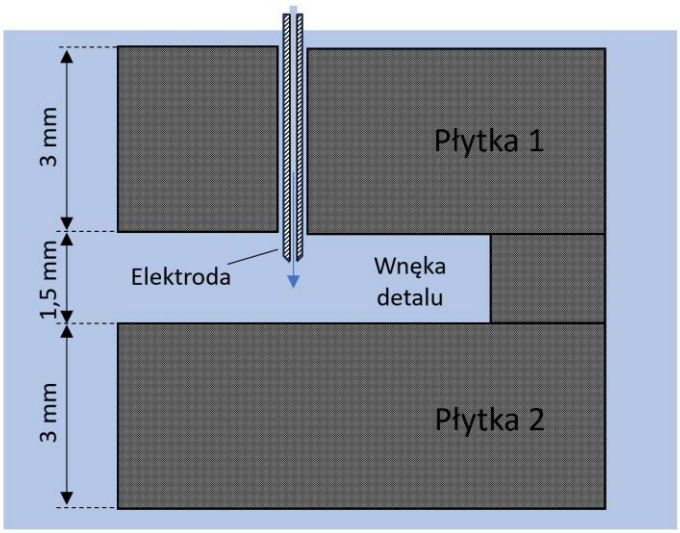
do urządzenia przez jednostkę filtracyjno- pompową. W urządzeniu zabudowany jest filtr 5µm. System filtrujący podaje do maszyny wodę, w której zanurzana jest część w czasie obróbki oraz zasila wrzeciono elektrody z pomocą pompy wysokiego ciśnienia. Układ sterujący drążarki rejestruje zmiany średniego napięcia i natężenia prądu co 6ms.

Testy drążenia wykonywane były na płytkach wykonanych ze stopu niklowo chromowego Inconel 718 o wymiarach 25x50x3mm. Materiał wybrany został ze względu na dobrą dostępność oraz szerokie wykorzystanie w technologiach silników lotniczych i badaniach procesu drążenia elektroerozyjnego.

Tabela 1 Skład chemiczny stopu Inconel 718

Ni [%]	Cr [%]	Nb [%]	Mo [%]	Ti [%]	Al. [%]	Co [%]	Cu [%]	C [%]	Si/Mn [%]	Fe [%]
58	21.5	3.6	9	0.4	0.4	1	0.3	0.1	0.35	5

Drążenia testowe wykonywane były na specjalnie przygotowanych zestawach składających się z dwóch płytek Inconel 718 z pustą przestrzenią pomiędzy, jak na poniższym rysunku 22.



Rysunek 22 Zestaw płytek testowych Inconel 718

Konstrukcja taka ma za zadanie symulować budowę części silnika. Przestrzeń pomiędzy płytkami odwzorowuje kanał wewnętrzy elementów turbiny, przez który normalnie tłoczony jest gaz o niskiej temperaturze tworzący film chłodzący na powierzchni części. Druga płytka w założeniach miała pełnić także rolę pomocniczą dla etapu identyfikacji momentu przebicia ścianki.

Otwory wykonywano standardowymi dla procesu produkcyjnego elektrodami mosiężnymi o średnicy 0.3mm. Aby zapewnić powtarzalność procesu, próbka statystyczna z każdej dostawy elektrod poddawana jest badaniom składu materiałowego. Skład chemiczny podano w poniższej tabeli.

Tabela 2 Skład chemiczny typowej elektrody mosiężnej

Składnik	Energia [keV]	% Masy [%]	Błąd [%]
Cu	8,04	68,67	1,53
Zn	8,63	31,33	1,96
Total		100,00	

4.1. System pomiarowy procesu FHD

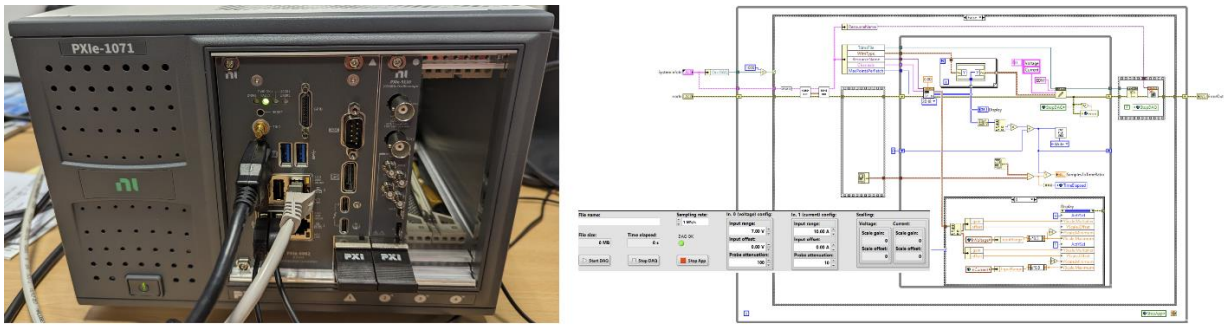
W celu poznania charakteru procesu drążenia otworów wentylacyjnych w procesie FHD przeprowadzono serię drążeń eksperymentalnych. Rejestrowane były parametry mierzone przez elektrodrążarkę oraz wysokoczęstotliwościowe przebiegi prądu oraz napięcia z pomocą zewnętrznych czujników. Dane z pomiarów posłużyły do uzyskania charakterystyki procesu, a w kolejnym etapie odpowiednio przygotowane dane zostały zastosowane do procesu wytrenowania modeli do wykrywania przebicia oraz wykrywania anomalii.

W celu rejestracji przebiegów prądu i napięcia zbudowano system akwizycji danych oparty o platformę sprzętową NI PXI i środowisko programistyczne LabVIEW.

W skład systemu pomiarowego wchodziły:

- NI PXIe-1071 - 4 slotowa obudowa PXI express, przepustowość do 3GB/s
- NI PXIe-8862 – kontroler PXI, procesor 8-rdzeniowy Intel Core i7-11850HE, 4.70 GHz, 16 GB DDR4-3200 MHz, 8.0 GT/s
- NI PXIe-5110 – 2 kanałowa karta oscyloskopowa, pasmo przepustowe 100 MHz, częstotliwość próbkowania do 1GS/s, rozdzielczość wejściowa 8 bit
- PXIe-6363 – karta wejściowa PXI, 32 wejścia napięciowe (rozdzielczość 16 bit, częstotliwość próbkowania do 2MS/s), 4 wyjścia analogowe ((rozdzielczość 16 bit, częstotliwość odświeżania do 2.86MS/s), 48 wejść/wyjść cyfrowych TTL

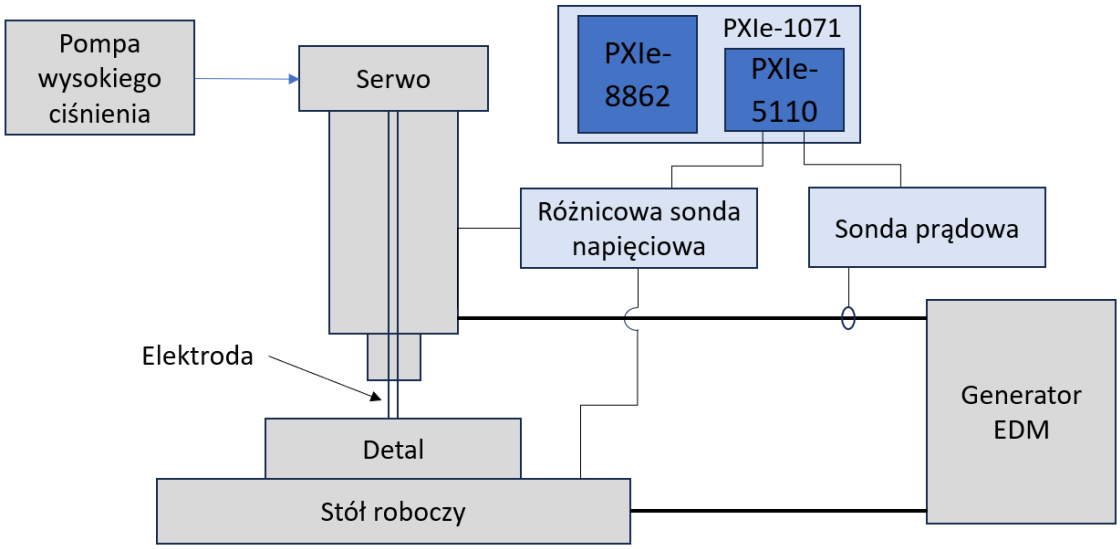
- Teledyne AP031 - różnicowa sonda napięciowa, pasmo przepustowe do 25 MHz
- HIOKI 3274 - Sonda prądowa, pasmo przepustowe do 10 MHz



Rysunek 23 Sterownik NI PXI wykorzystany do rejestracji prądu i napięcia oraz fragment kodu źródłowego aplikacji akwizycji danych

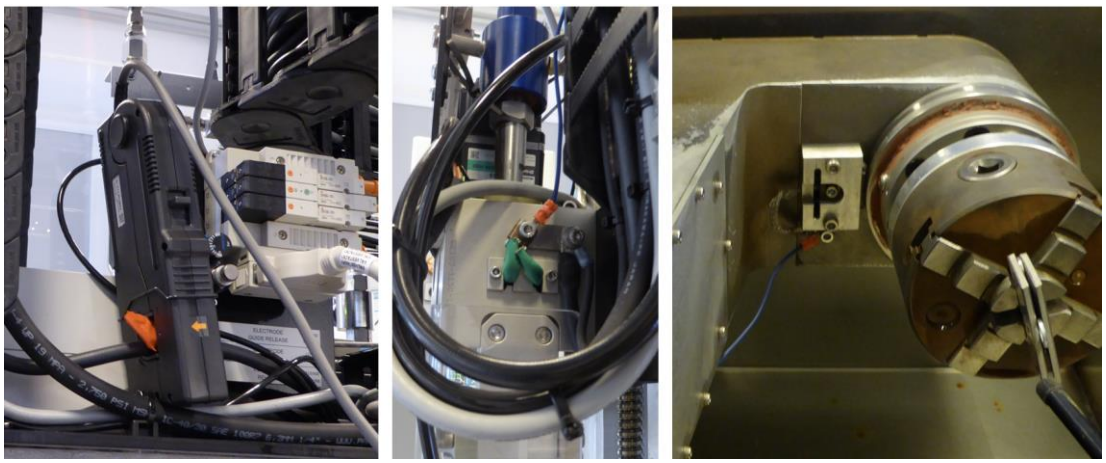
Na potrzeby badań zbudowana została dedykowana aplikacja akwizycji danych pozwalająca na wykorzystanie kart pomiarowych NI PXIe-5110 i PXIe-6363 oraz środowiska programistycznego LabVIEW. Aplikacja pozwala na łatwą konfigurację parametrów mierzonych kanałów (takich jak parametry skalowania) oraz wybór częstotliwości próbkowania w zakresie 100kS/s-10MS/s. Z każdego pomiaru generowany był oddzielny plik danych w formacie .tdms zawierający przebiegi prądu i napięcia z całego drążenia.

Pierwszą grupę pomiarów testowych wykonano z wykorzystaniem karty NI PXIe-5110. Układ pomiarowy pozwalał na mierzenie prądu oraz napięcia ze zmienną częstotliwością próbkowania sięgającą 10 MS/s. Podejście to miało na celu określenie optymalnej częstotliwości pomiarowej (rozdział 4.3). Ideowy schemat układu pomiarowego pokazano na rysunku 24.



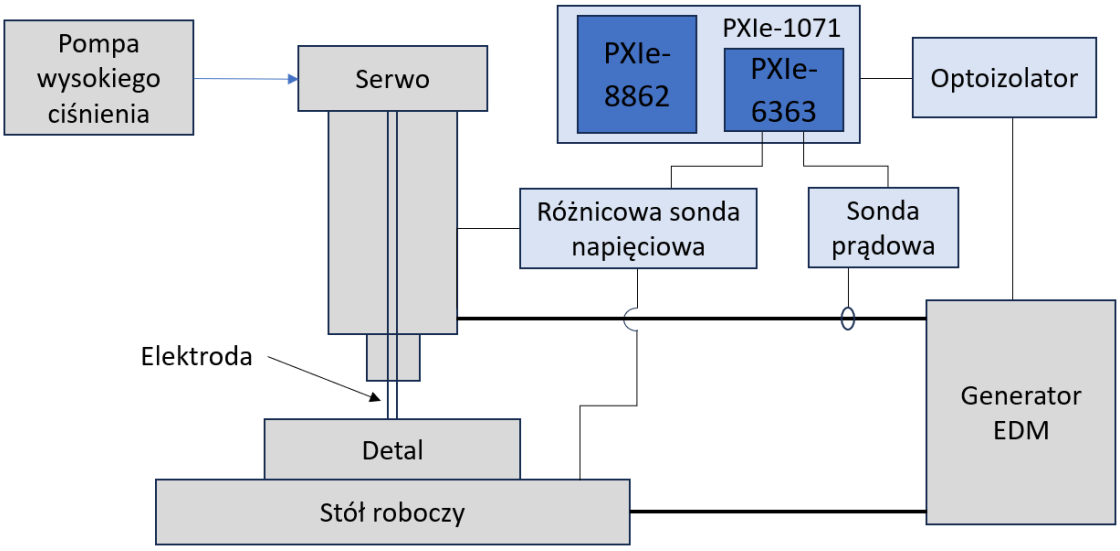
Rysunek 24 Ideowy schemat testowego układu pomiarowego 1 (karta pomiarowa PXIe-5110)

Pozostałe parametry procesu rejestrowane były przez sterownik elektrodrążarki. System sterowania maszyny zapisywał parametry pracy z częstotliwością próbkowania 160 S/s w plikach .csv, tworząc nowy plik dla każdego drążonego otworu. Maszyna rejestruje wartości takie jak: położenie wrzeciona, prędkość posuwu osi Z, napięcie szczeliny, temperatura wody, konduktancja wody, zmiana prędkości wrzeciona, zmiana napięcia szczeliny, ciśnienie wody, przemieszczenie w czasie wiercenia, prąd szczeliny, uśredniony posuw, przepływ dielektryka przez elektrodę.



Rysunek 25 Punkty pomiarowe na maszynie FHD

W kolejnych próbach drążenia, po określeniu minimalnej wymaganej częstotliwości pomiarowej, zmodyfikowano układ pomiarowy do zastosowania karty pomiarowej PXIe-6363, jak pokazuje poniższy rysunek 26.



Rysunek 26 Ideowy schemat testowego układu pomiarowego 2 (karta pomiarowa PXIe-6363)

Zastosowanie innego podejścia pomiarowego podyktowane było zidentyfikowanymi w analizach danych problemami z synchronizacją zegarów maszyny i układu pomiarowego PXI (szerzej opisywanych w rozdziale 4.6). Układ pozwalał na dodatkowe zmierzenie sterowniczego sygnału prostokątnego TTL podawanego na tranzystorowy układ generatora iskier EDM. Sygnał prostokątny jest odpowiedzialny za określanie okresu T_{on} (sygnał sterowniczy na poziomie logicznym 1) i T_{off} (sygnał na poziomie logicznym 0). W celu zabezpieczenia i zapewnienia izolacji galwanicznej układu pomiarowego, wzbogacono go o odpowiedni układ optoizolacyjny. Dodatkowo karta pozwoliła na określenie ilości informacji traconych ze względu na ograniczoną rozdzielczość bitową karty PXIe-5110.

4.2. Drążenia testowe

W celu zebrania danych eksperymentalnych dokonano szeregu drążeń testowych na próbnych płytkach z Inconel 718, stosując trzy zestawy parametrów procesu obrazujących różne warianty wierceń. Zestawienie parametrów optymalnych (zestaw #1), agresywnych (zestaw #2) i błędnych (zestaw #3) drążenia przedstawiono w poniższej tabeli (Tabela 3).

Tabela 3 Tabela parametrów procesu drążenia

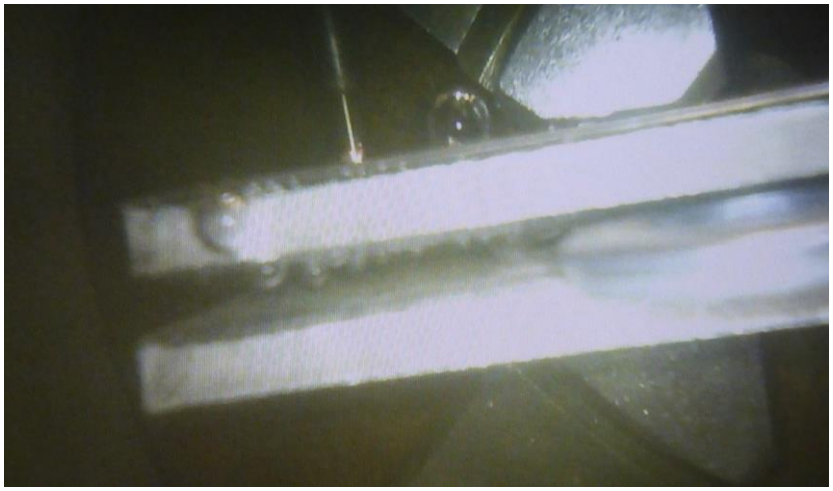
Zestaw parametrów	Ton [µs]	Toff [µs]	Ip [A]	Napięcie szczeliny [V]	Pojemność	Obroty elektrody [obr/min]	Ciśnienie płukania [bar]
Zestaw parametrów #1 (optymalne)	70	30	10	60	8	690	29
Zestaw parametrów #2 (agresywne)	30	30	10	40	8	690	29
Zestaw parametrów #3 (błędne)	50	10	15	10	2	690	29

W rozprawie zastosowano uogólnione podejście wykorzystujące parametry procesu zbliżone do parametrów produkcyjnych. Podejście takie miało na celu udowodnienie prawidłowości działania opracowywanych innowacyjnych metod kontroli procesu dla wybranych przypadków występujących w procesie FHD.

Badania procesu drążenia FHD wykazały, że wyniki procesu drążenia zależą od wielu czynników, także czynników zmiennych w czasie procesu. Przykładem może być np. długość elektrody, która ulega skracaniu z każdym kolejnym drążeniem. Zmiana długości elektrody, szczególnie dla otworów o małych średnicach, może wpływać na stabilizację jej położenia przy wprowadzaniu w ruch obrotowy oraz ma wpływ na warunki płukania (skrócenie długości elektrody zmniejsza opory przepływy dielektryka zwiększając jego wydatek). Czynniki te wprowadzają istotną zmienność w procesie. Jako że celem niniejszej pracy nie jest analizowanie wpływu i optymalizacja parametrów samego procesu, a raczej opracowanie uniwersalnych rozwiązań analizy i kontroli procesu, wprowadzono podejście mające na celu niwelację tych czynników. Zestawy danych uczących tworzone były z dużej ilości danych zarejestrowanych w różnych drążeniach. Każdy kolejny otwór wykonywany był elektrodą o zmienionej długości, od długości maksymalnej do minimalnej standardowo używanej w procesie produkcyjnym. Pomiędzy grupami drążeń, kilkukrotnie wykonywana była zmiana elektrod. Miało to na celu uwzględnienie w populacji uczącej danych z otworów wykonywanych w wielu następujących po sobie drążeniach. Podejście takie zapewniło uwzględnienie zmian długości elektrody w czasie drążenia oraz powinno wprowadzań generalizację czynnika niepewności wynikającego z jakości wykonania samej elektrody. Podejście takie dobrze symuluje faktyczny przebieg procesu w czasie produkcji części silników odrzutowych.

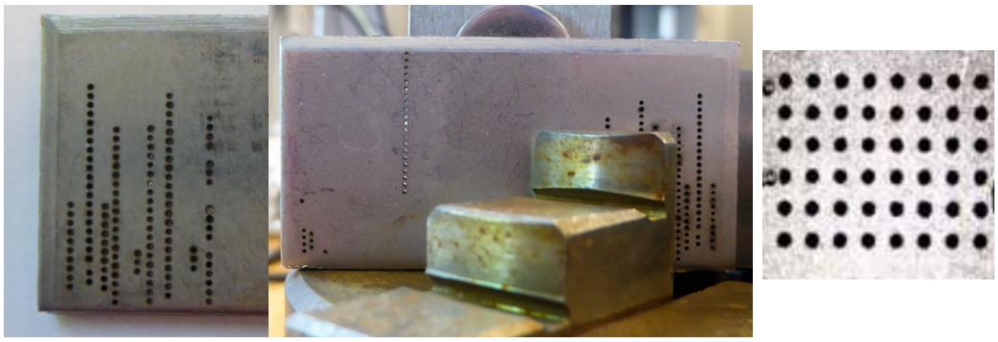
Wiercenia z wykorzystaniem układu pomiarowego 1 wykonywane były z elektrodą ustawioną prostopadle do wierconej powierzchni. Na rysunku 27 przedstawiono zdjęcie

procesu wiercenia z kamery zanurzonej w dielektryku wewnątrz przestrzeni roboczej elektrodrążarki.



Rysunek 27 Zdjęcie procesu drążenia z elektrodą ustawioną prostopadle do drążonej powierzchni

Każdy z otworów wiercony był w nowej pozycji na detalu. Proces polegał na całkowitym przewierceniu pierwszej płytki i kończony był, kiedy elektroda zaczynała dokonywać wiercenia w płytce drugiej. Podejście to, miało za zadanie ułatwienie identyfikacji momentu całkowitego zakończania drążenia otworu w pierwszej płytce testowej. Korelacja momentu wykrycia rozpoczęcia procesu wiercenia w drugiej płytce testowej z danymi o przemieszczeniu osi Z wrzeciona roboczego powinna wstępnie określać moment w jakim elektroda kończy przebijać się przez pierwszą ściankę detalu testowego. Na poniższym zdjęciu przedstawiono przykładowe płytki po procesie wiercenia.



Rysunek 28 Przykładowe płatki Incolen718 po wierszowym drążeniu otworów w procesie FHD

Otwory wykonywane były w procesie tzw. wierszowania. Kolejne drążenia oznaczane były przez numer wiersza (np. R1, R5) oraz numer kolejnego otworu w wierszu (np. H1, H7).

W pierwszej turze pomiarów dla każdego z dwóch poprawnych zestawów parametrów procesu wykonano po 10 odrębnych otworów z różnymi częstotliwościami próbkowania sygnałów napięcia i prądu, kolejno 500kS/s, 1MS/s, 2MS/s, 5MS/s oraz 10MS/s przy wykorzystaniu karty PXIe-5110. Zestawienie pomiarów testowych przedstawiono w poniższej tabeli (Tabela 4).

Tabela 4 Zestawienie pomiarów testowych, otwory przelotowe elektroda prostopadła do płytek testowych

Częstotliwość próbkowania [MS/s]	Ilość otworów [szt.]	Rząd otworów	Oznaczenie Otworów	Zestaw parametrów	Orientacja elektrody
10	10	R1	H1-H10	#1	prostopadła
5	10	R2	H1-H10	#1	prostopadła
2	10	R3	H1-H12	#1	prostopadła
1	10	R4	H1-H10	#1	prostopadła
0,5	10	R5	H1-H11	#1	prostopadła
10	10	R6	H1-H10	#2	prostopadła
5	10	R7	H1-H10	#2	prostopadła
2	10	R8	H1-H10	#2	prostopadła
1	10	R9	H1-H10	#2	prostopadła
0,5	10	R10	H1-H10	#2	prostopadła

Dla każdego wierconego otworu rejestrowany był także zestaw parametrów sterownika maszyny, przede wszystkim: bezwzględne i względne położenie wrzeciona w osi Z, napięcie szczeliny, stosunek prędkości (zmiana prędkości wrzeciona w czasie), stosunek napięcia (zmiana napięcia EDM w czasie), ciśnienie płukania elektrody, średni posuw wrzeciona, przepływ dielektryka płukania elektrody.

Zarejestrowane dane o częstotliwości próbkowania innej niż określona docelowa częstotliwość, po procesie resamplowania (opisanym w rozdziale 4.6) posłużyły jako dane uczące modeli.

Po określeniu wymaganej częstotliwości próbkowania (patrz rozdział 4.3) i wstępnych analizach danych, wykonano serię kolejnych prób z wykorzystaniem karty PXIe-6363 (układ pomiarowy 2). Pomiary wykonywano z określoną na podstawie analizy danych z układu pomiarowego 1 częstotliwością próbkowania 500 kS/s. Oprócz napięcia oraz prądu do mierzonych wartości dodano sygnał nośny odpowiedzialny za generowanie czasów T_{on} i T_{off} ze względu na konieczność potwierdzenia synchronizacji czasowej przebiegów (patrz rozdział 4.6). Mierzony sygnał był cyfrowym sygnałem prostokątnym w standardzie TTL o określonym

stopniu wypełnienia. Sygnał ten trafiał w maszynie na układ tranzystorowy odpowiedzialny bezpośrednio za generację czasów impulsu i czasu bezimpulsowego. Tak jak w poprzednich próbach rejestrowane również były dane zapisywane przez maszynę MKG 1000.

Kolejnej serii pomiarów powtórzono pomiary z elektrodą prostopadłą do płytek oraz wykonano szereg wierceń z płytką ustawioną pod kątem 45° względem elektrody dla obu zestawów parametrów. Zestawienie drążeń testowych pokazuje poniższa tabela.

Tabela 5 Zestawienie pomiarów testowych z użyciem PXIe-6363, elektroda prostopadła i ustawiona pod kątem do płytek testowych

Częstotliwość próbkowania [MS/s]	Ilość otworów [szt.]	Rząd otworów	Oznaczenie Otworów	Zestaw parametrów	Orientacja elektrody
0,5	60	R11-R13	H1-H20	#1	prostopadła
0,5	60	R14-R16	H1-H20	#2	prostopadła
0,5	60	R17-19	H1-H20	#1	kątowa
0,5	60	R20-R22	H1-H20	#2	kątowa

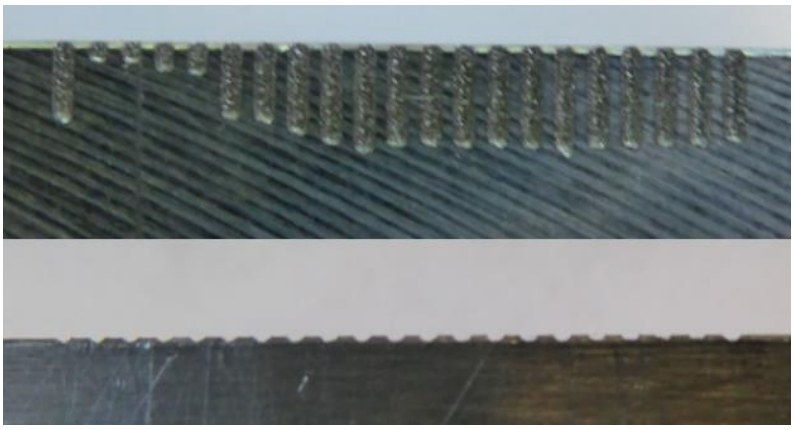
Celem było zebranie ilości danych, które powinny pozwolić na szkolenie sieci neuronowych o dobrych możliwościach generalizacji całej populacji możliwych do wystąpienia stanów drążenia oraz minimalizacja opisywanego wcześniej wpływu zmiennych warunków procesu, jak np. długość elektrody. W pracy z premedytacją nie wykonywano bardzo dużej ilości otworów. Założeniem pracy jest udowodnienie, że już relatywnie nieduża ilość otworów testowych, porównywalna z nawierceniem pełnych zestawów otworów wentylacyjnych w zaledwie kilku sztukach części silnika będzie wystarczające do uruchomienia tworzonych w rozprawie narzędzi w warunkach produkcyjnych.

Następnie wykonano serię pomiarów mającą na celu symulację potencjalnych usterek mogących wystąpić podczas wierceń. Zasymulowano wiercenia z całkowicie błędnymi ustawieniami parametrów maszyny (parametry #3). Sytuacja taka może symulować uszkodzenie któregoś z elementów kontrolnych maszyny lub przypadkowe używanie niepoprawnego zestawu parametrów w warunkach produkcyjnych. Przeprowadzono także serię wierceń symulujących obserwowane podczas wcześniejszej kampanii testowej błędy mogące wystąpić w warunkach produkcyjnych: wiercenia z ograniczonym przepływem dielektryka przez elektrodę, wiercenia wykonywane wygiętą elektrodą oraz zasymulowano zjawisko scarfingu (patrz rozdział 2.3.2).

Tabela 6 Zestawienie pomiarów testowych symulujących usterki FHD

Częstotliwość próbkowania [MS/s]	Ilość otworów [szt.]	Rząd otworów	Oznaczenie Otworów	Zestaw parametrów	Orientacja elektrody	Usterka
0,5	20	R23	H1-H20	#3	prostopadła	Błędne parametry
0,5	20	R24	H1-H20	#1	prostopadła	Ograniczony przepływ
0,5	20	R25	H1-H20	#1	prostopadła	Wygięta elektroda
0,5	20	R26	H1-H20	#1	prostopadła	scarfing

Na rysunku 29 przedstawiano wygląd fragmentu materiału umieszczonego pomiędzy płytkami 1 i 2 po zasymulowaniu na nim zjawiska scarfingu.

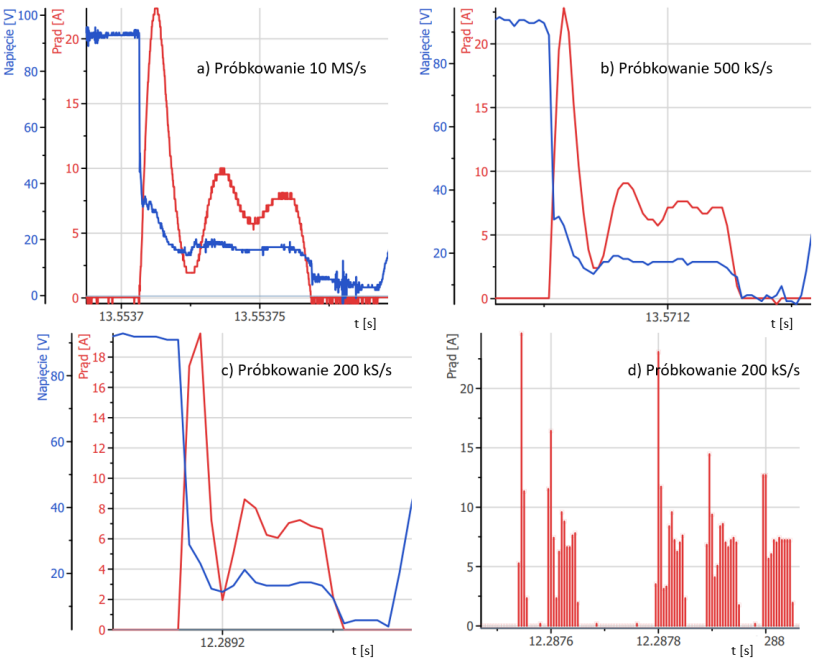


Rysunek 29 Wynik drążeń symulujących proces scarfingu

4.3. Określenie minimalnej wymaganej częstotliwości próbkowania sygnałów

Stosowanie różnych częstotliwości pomiarowych miało na celu określenie minimalnych wymogów docelowego systemu pomiarowego, które będą pozwalały na dostatecznie dokładne identyfikowanie cech charakterystycznych mierzonych wyładowań. Z punktu widzenia docelowego systemu istotne jest ustalenie minimalnej wymaganej częstotliwości próbkowania. Zmniejszenie częstotliwości próbkowania zmniejsza ilość koniecznych do przetworzenia danych, co przekłada się na optymalizację obciążenia obliczeniowego zastosowanego algorytmu klasyfikacji etapu drążenia i wykrywania anomalii. Należy jednak mieć na uwadze, że zastosowanie zbyt małej częstotliwości próbkowania będzie prowadziło do potencjalnej utraty istotnych informacji jak np. brak możliwości niezawodnego rejestrowania prawdziwej maksymalnej amplitudy poszczególnych wartości pików prądowych.

Analiza danych pomiarowych wykazała, że sygnały napięcia oraz prądu powinny być próbkowane z częstotliwością przynajmniej 500kS/s. Na poniższym rysunku (Rysunek 30) przedstawiono przykładowe przebiegi prądu (kolor czerwony) i napięcia (kolor niebieski) jednego impulsu EDM, zarejestrowane przy próbkowaniu 10 MS/s (Rysunek 30 a)), 500 kS/s (Rysunek 30 b)), oraz 200 kS/s (Rysunek 30 c)).



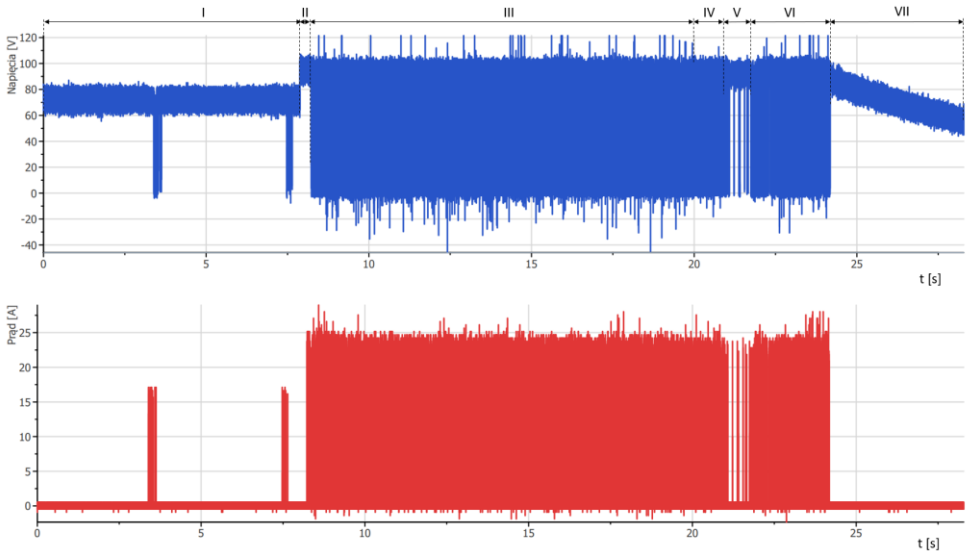
Rysunek 30 Porównanie sygnałów napięcia i prądu dla różnych częstotliwości próbkowania, a) przebiegi 10 MS/s, b) przebiegi 500 kS/s, c) przebiegi 200 kS/s, d) Wizualizacja dyskretnych wartości dla próbkowania 200 kS/s

Jak widać, sygnały impulsu próbkowane z częstotliwością 500kS/s nadal prawidłowo oddają charakter przebiegu. Poprawnie odwzorowane zostają zarówno kształt przebiegu jak i wartości amplitud sygnałów. W celu sprawdzenia poprawności założenia o poprawnym odwzorowaniu amplitudy impulsów prądowych, pliki zarejestrowane z maksymalną częstotliwością zostały resamplowane (stosowany proces resamplowania szerzej opisany zostanie rozdziale 4.6) do innych testowanych częstotliwości. Następnie porównywano wartości amplitud losowo wybranych impulsów. Dla częstotliwości 500kS/s maksymalna różnica wartości prądów szczytowych wynosiła 1,6%. W wypadku rejestracji sygnału z częstotliwością próbkowania 200kS/s obserwowana jest znaczna degradacja informacji o procesie. Ogólny kształt sygnałów zostaje zachowany, natomiast tracona jest informacja o prawdziwej maksymalnej amplitudzie prądu i napięcia. Dla tej częstotliwości błąd odtworzenia maksymalnej amplitudy prądu sięgał 22%. Jak pokazano na czwartym wykresie (Rysunek 30 d)), próbkowanie 200kS/s zapewnia rejestrację pojedynczej próbki w chwili

osiągania maksimum prądu płynącego w obwodzie. Powoduje to, że ze względu na nieuniknione różnice w rejestracji poszczególnych drążeń wynikające np. z przesunięcia czasowego startu rejestracji sygnałów czy różnic w ustawieniach czasów T_{on} i T_{off} procesu, minimalną dopuszczalną częstotliwością próbkowania sygnałów powinno być 500kS/s.

4.4. Charakteryzacja przebiegów czasowych

Dokonano analizy czasowych przebiegów prądu i napięcia w celu identyfikacji charakterystycznych etapów procesu drążenia. Rysunek 31 przedstawia przebieg napięcia i prądu całości przykładowego drążenia testowego.



Rysunek 31 Napięcie i prąd drążenia otworu z identyfikacją etapów procesu (otwór R2 H7)

W każdym z drążeń testowych można zidentyfikować kilka charakterystycznych etapów (etapy od I do VI).

- Etap I jest przygotowaniem do drążenia. W tej części procesu napięcie robocze zostaje zredukowane, a elektrodrążarka dokonuje podjazdu do powierzchni drążonego detalu z zadaną prędkością w celu wykrycia położenia powierzchni elementu obrabianego. W momencie zmniejszenia szczeliny pomiędzy elektrodą, a detalem do odległości pozwalającej na jonizację dielektryka następuje inicjacja generowania iskier. Obrazuje to pojawianie się impulsów prądowych oraz skorelowane z nimi spadki napięcia. Po wykryciu impulsów przez sterownik maszyny następuje wycofanie wrzeciona i powtórzenie procesu po raz drugi.

- Etap II – Po namierzeniu położenia powierzchni detalu następuje etap drążenia. Napięcie zostaje podniesione do ustawionego napięcia roboczego. Załączany jest posuw wrzeciona w kierunku detalu z zadaną prędkością.

- Etap III – po dojechaniu do powierzchni detalu następuje faktyczny proces EDM. Elektroda przesuwana jest w osi Z obrabiarki. Generowana jest seria wyładowań elektrycznych, które systematycznie erodują przedmiot obrabiany i elektrodę pozwalając na pogłębianie drążonego otworu. Proces prowadzony jest w sposób ciągły, aż do stworzenia przelotowego otworu w obrabianym detalu.

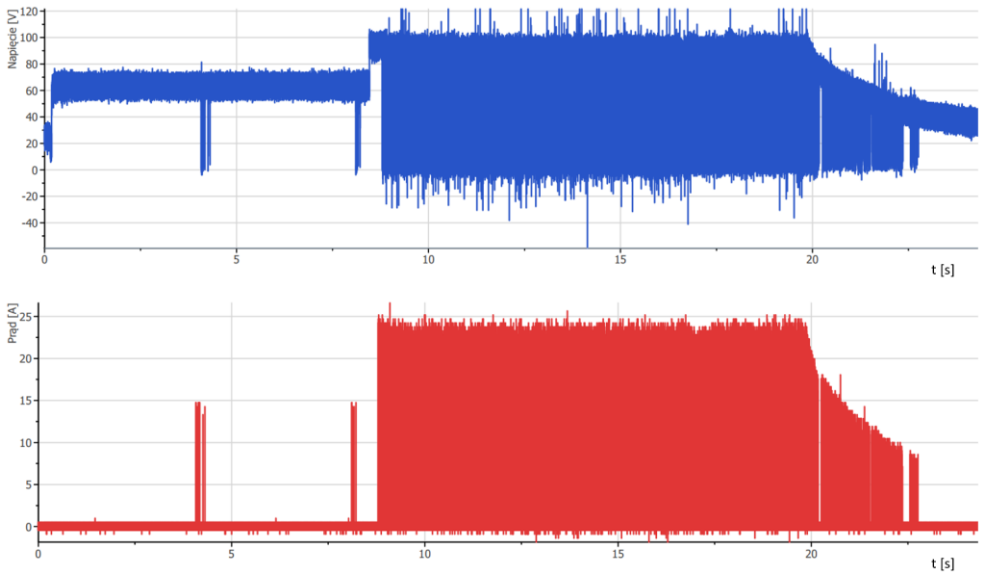
- Etap IV – elektroda dochodzi do przeciwległej ścianki detalu i następuje jej przebicie (tzw. przełom). Elektroda dochodząc do ścianki detalu tworzy najpierw niewielki otwór, który później zostaje rozszerzony, aby pozwolić na całkowite przejście elektrody przez wytworzony otwór i zakończenie drążenia w pierwszej płytce. Ze względu na stopniowe formowanie otworu połączone z równoczesnym pogorszeniem warunków płukania, wykrycie dokładnego momentu zakończenia otworu jest wyjątkowo trudne. Po całkowitym przebiciu elektroda jest dalej przemieszczana wzdłuż osi Z w głąb wnęki pomiędzy płytkami.

- Etap V – elektroda przemieszcza się przez wnękę pomiędzy płytkami. W tym etapie nadal może zachodzić proces tworzenia iskier w szczelinie bocznej między elektrodą, a przedmiotem obrabianym. Nie zachodzi za to tworzenie iskier na czole elektrody, szczelina boczna w płytce 1 nie jest efektywnie płukana.

- Etap VI – elektroda dochodzi do drugiej płytki testowej i rozpoczynane jest jej drążenie w procesie takim samym jak drążenie etapu III. Symuluje to opisywany wcześniej błąd procesu nazywany jako „Overdrilling” lub „backwall strike”. Drążenie zatrzymywane jest po upływie ustawionego czasu mierzonego od startu etapu III. Czas został ustawiony w taki sposób, aby zagwarantować rozpoczęcie drążenia w drugiej płytce podczas każdego z drążeń testowych.

- Etap VII – po upływie ustawionego czasu następuje wycofanie wrzeciona roboczego i zakończenie drążenia otworu. Detal obrabiany przemieszczany jest w nowe położenie w przestrzeni roboczej obrabiarki przygotowując proces do drążenia kolejnego otworu.

Na opisywanym powyżej rysunku przebiegu sygnałów zaobserwować można czasowy zanik generowania impulsów w etapie V, który kończy się dojazdem do drugiej płytki i ponownym rozpoczęciem generowania impulsów. Niestety, sytuacja pokazana na rysunku 31 jest sytuacją wyjątkową, która w warunkach produkcyjnych zachodzi bardzo rzadko. Łatwo identyfikowalna przerwa w generacji impulsów widoczna na przebiegach czasowych prądu i napięcia wystąpiła tylko w kilku spośród wszystkich wywierconych otworów testowych. Nawet wtedy zaprzestanie generacji impulsów nie musi być jednoznaczne z prawdziwym momentem zakończenia formowania otworu. W wypadku otworów wierconych z drugim zestawem parametrów maszyny (parametry „agresywne"), widoczna przerwa w generacji impulsów nie wystąpiła ani razu. W znacznej większości otworów eksperymentalnych typowe przebiegi czasowe wyglądały jak poniżej (Rysunek 32).

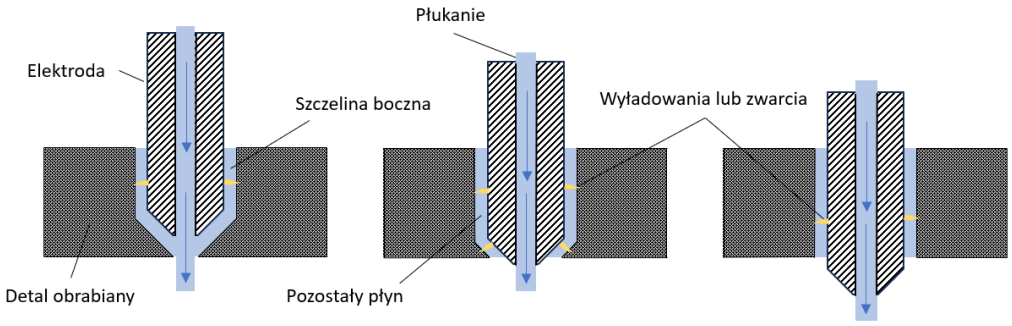


Rysunek 32 Typowy czasowy przebieg napięcia i prądu procesu (otwór R10 H9)

Analizując dane z maszyny i obserwacje przebiegu procesu wiemy, że przebicie w powyższym procesie nastąpiło gdzieś pomiędzy 15, a 20 sekundą procesu. Jak widać analizując same przebiegi czasowe bardzo trudno jest określić moment przebicia.

Dzieje się tak ponieważ, w fazie przełomu płyn roboczy przestaje efektywnie usuwać zanieczyszczenia. Dielektryk podawany przez środek elektrody po przełomie zostaje wpompowywany do wnęki detalu zamiast, tak jak wcześniej, omywać dolną szczelinę i szczelinę boczną między elektrodą, a detalem. Powoduje to nagromadzenie się dużej ilości zanieczyszczeń w obszarze bocznej szczeliny. Nagromadzone zanieczyszczenia, mogą powodować wtórne wyładowania i zwarcia pomiędzy elektrodą, a bocznymi ściankami

detalu, prowadząc do jeszcze większego nagromadzenia zanieczyszczeń w tej przestrzeni. Dodatkowo, dielektryk płynący pod dużym ciśnieniem przez szczelinę boczną w czasie drążenia może zapewniać stabilizację pozycji elektrody. Zanik przepływy może przyczyniać się do pogorszenia pozycjonowania, szczególnie dla długich elektrod. Przed zakończeniem wiercenia, wyładowania i zwarcia mogą wystąpić zarówno w szczelinie bocznej, jak i dolnej. Po zakończeniu otworu, jednakże te impulsy mogą występować tylko w bocznej szczelinie, niezależnie od położenia elektrody. Rysunek 33 przedstawia ilustracje różnych etapów fazy tworzenia przełomu.



Rysunek 33 Przebieg fazy przełomu w procesie FHD

Choć impulsy w bocznej i dolnej szczelinie mogą być trudno odróżnialne, spodziewa się, że ilość wystąpień różnych typów wyładowań będzie się zmieniała przed i po zakończeniu formowania otworu. Ze względu na to, że wyładowania i zwarcia przestają generować się w dolnym luźnie, spodziewa się, że ilość wyładowań normalnych, łukowych i zwarć będą maleć, a ilość wystąpień otwartego obwodu będzie wzrastać [64-68].

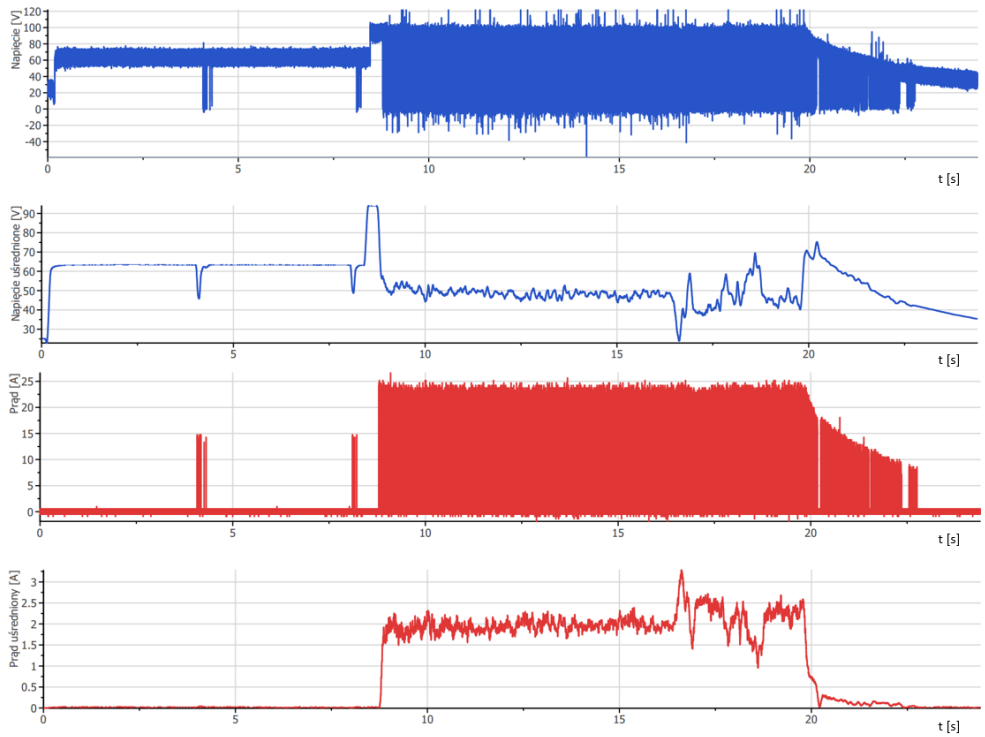
W celu wstępnego potwierdzenia założenia o zmianie częstotliwości występowania różnych rodzajów impulsów w fazie przełomu oraz określenia potencjalnego miejsca przełomu i zakończenia wiercenia otworu, przebiegi czasowe zostały poddane wygładzaniu poprzez uśrednianie wartości prądu i napięcia. Jeśli założenie o zwiększonej ilości obwodów otwartych po zakończeniu otworu jest prawdziwa, zależność ta powinna być widoczna także w uśrednionych wartościach prądu i napięcia.

Zastosowana funkcja wygładzania zastępuje każdą pojedynczą wartość kanału nieobciążoną wartością średniej arytmetycznej określonej liczby N sąsiednich wartości. Wyjściowa ilość próbek danego kanału pomiarowego pozostaje taka sama. Funkcja wygładzania nie wykonuje redukcji ilości punktów pomiarowych.

Wartość średnia oblicza wygładzanie z $2N + 1$ wartości według następującej formuły, gdzie m to długość kanału.

$$f(y_i) = \frac{1}{2N+1} \sum_{k=i-N}^{k=i+N} y_k\,, \qquad m - N > i > N \tag{15}$$

Na początku i na końcu kanału, odległość wartości od zewnętrznej strony jest mniejsza niż jednostronna szerokość wygładzania N, a wygładzanie jest obliczane odchodząc od formuły przy mniej niż 2N+1 wartościach.



Rysunek 34 Wygładzone przebiegi napięcia i prądu otworu R10 H9

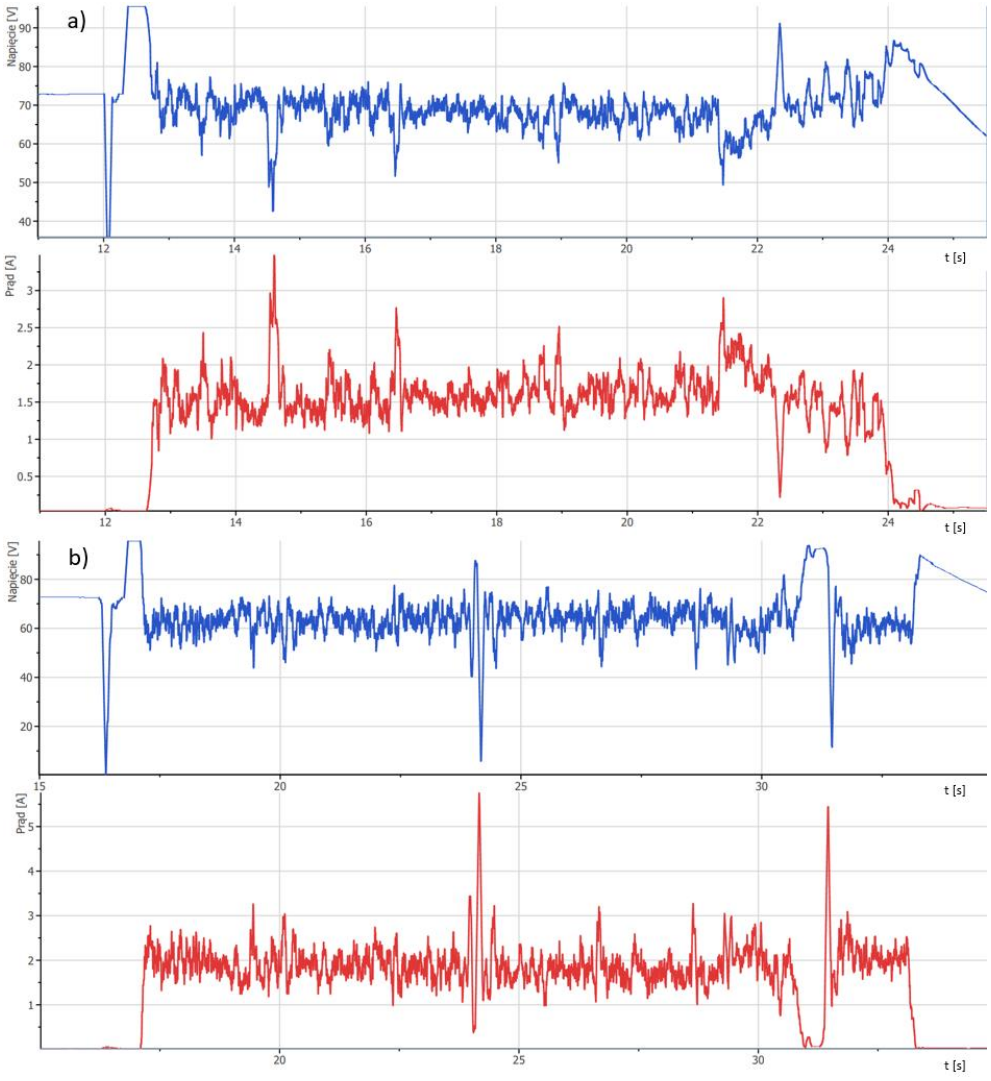
Rysunek 34 pokazuje przykładowe porównanie sygnałów czasowych z sygnałami wygładzonymi. W przykładzie próbkowanym z częstotliwością 500 kS/s, zastosowano uśrednianie z 20000 sąsiednich próbek, co przekłada się na uśrednianie wartości mierzonych z okresu 80ms. Przetestowano wygładzanie przebiegów z różnych okresów. Stosowanie uśredniania z mniejszych okresów skutkowało przebiegami o dużej zmienności, co utrudniało jednoznaczne określanie momentów, w których następowała zmiana trendu. Dłuższe okresy uśredniania nadal pokazują trend, natomiast tracona była część informacji o zmienności sygnału. Ponieważ celem jest jak najszybsze wykrywanie przebicia otworu, istotne jest zachowanie dynamiki zmian. Jak pokazuje powyższy rysunek, pomimo trudności

w zaobserwowaniu wystąpienia przełomu w procesie wiercenia otworu, analizując przebiegi uśrednione można zauważyć zmianę w średnich wartościach zarówno prądu jak i napięcia w okolicach 17 sekundy procesu. Sugerować to może poprawność założenia o zmianie charakteru i ilości impulsów generowanych w trakcie przełomu [64, 65, 69]. Co ciekawe, pierwszą zmianą jaką zaobserwowano jest wzrost średniej wartości prądu połączony ze spadkiem średniego napięcia. Dzieje się tak najprawdopodobniej ze względu na opisywane wcześniej pogorszenie warunków płukania i pozycjonowania elektrody. Ta sytuacja przekładać się może na zwiększoną ilość wyładowań różnego typy, pomimo faktu zmniejszania powierzchni (utworzenie przełomu) pomiędzy którymi mogą powstawać iskry. Po czasowym zwiększeniu wartości wygładzonej prądu zaobserwowano spadek tej wartości. Spadek wartości prądu skorelować należy ze zmniejszeniem ilości generowanych wyładowań w czasie kończenia otworu i podczas przemieszczania elektrody przez wewnętrzną wnękę detalu. Czasowy spadek wartości charakteryzuje się niestety dużą zmiennością pomiędzy drążeniami kolejnych otworów. Ostatecznie obserwujemy ponowny wzrost wartości prądu do średniej wartości podobnej do wartości podczas wiercenia pierwszej płytki. Oznacza to dojazd elektrody do płytki 2 i ponowne rozpoczęcie wiercenia.

Proces wiercenie charakteryzuje się bardzo dużą zmiennością, także w analizowanych przebiegach wygładzonych. W części zarejestrowanych przebiegów można zaobserwować opisane powyżej trendy zmienności średnich napięć i prądów, natomiast sama zmienność procesu praktycznie uniemożliwia opieranie diagnostyki tylko na osiąganiu twardo ustalonych wartości progowych tych wartości.

Na rysunku 35 przedstawiono przebiegi wartości wygładzonych dla dwóch takich przypadków. W obu drążeniach już w czasie wiercenia wewnątrz płytki 1 (etap III procesu) występują wyraźne lokalne szczyty wartości prądu. W drążeniu otworu R5 H6 (Rysunek 35 b)) wystąpiły zarówno lokalne szczyty jak i doliny wartości prądu wygładzonego. Na dodatek, przy wierceniu tego otworu nie wystąpił bardzo wyraźny szczyt wartości prądu w momencie przełomu. Widoczny jest pewien trend wzrostowy wartości, natomiast szczyty wartości prądu w tym okresie są niższe niż szczyty obserwowane w czasie etapu III drążenia. Dalej zaobserwować dało się spadek średniej wartości prądu, najprawdopodobniej skorelowany z zakończeniem otworu. Potwierdza to hipotezę, że opieranie algorytmu wykrywania na stałych, z góry określonych wartościach progowych dla uśrednionych lub filtrowanych wielkości elektrycznych będzie wiązało się z potencjalnie błędnym

wskazaniem przebicia (błąd typu I). Błędy tego typu można by ograniczyć poprzez zastosowanie śledzenia trendu i podjęcie próby wykrycia przebicia przy dostatecznie długim utrzymaniu się kierunku zmian. Problemem tego podejścia jest nieuniknione wprowadzenia dodatkowego opóźnienia w procesie wykrywania. Ponadto, ze względu na dużą zmienność przebiegu procesu, potencjalnie różne ilości i typy generowanych wyładowań już podczas zasadniczego drążenia (etap III) mogą dawać takie same wartości wielkości uśrednionych i prowadzić do błędnych wniosków. Z tego względu zasadne jest podejście, które będzie brało pod uwagę także charakter pojedynczych wyładowań.

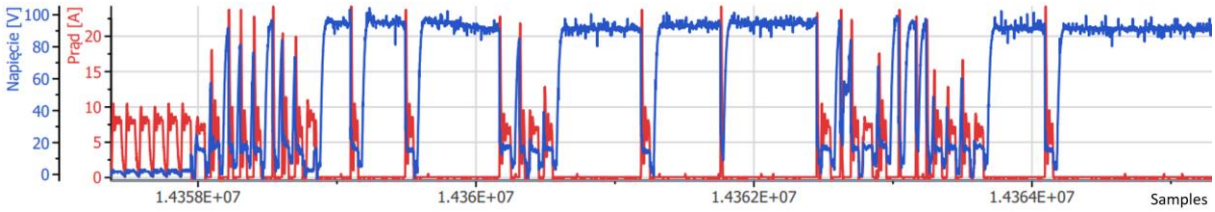


Rysunek 35 Duża zmienność wygładzonych przebiegów napięcia i prądu otworu R5 H1 (a) i R5 H6 (b)

4.5. Analiza charakteru impulsów EDM

Jak opisano w poprzednim rozdziale, analiza przebiegów czasowych w skali makro nie przynosi wystarczająco dokładnych informacji do pewnego określania etapu procesu. Postawiona natomiast została teza mówiąca o zmianie charakteru generowanych w tym czasie impulsów jak sugerują prace Li, Xia i Wanga [64, 65, 69]. W celu lepszego zrozumienia procesu oraz potwierdzenia założenia o podobieństwie typowych impulsów FHD do impulsów obserwowanych podczas innych procesów EDM (rozdział 3.2), dokonano szczegółowej analizy charakteru poszczególnych wyładowań zarejestrowanych podczas drążeń testowych.

Głębsza analiza przebiegów napięcia i prądu z drążeń, pokazuje wyjątkowo dużą zmienność tych sygnałów. Nawet w czasie optymalnego drążenia (parametry #1), w czasie wykonywania etapu III, zaobserwowano występowanie każdego z typowych impulsów procesu EDM z różną częstotliwością (impulsów normalnych, impulsów łukowych, obwodów otwartych i zwarć).



Rysunek 36 Zmienność przebiegów prądu i napięcia w czasie realizacji etapu III procesu FHD

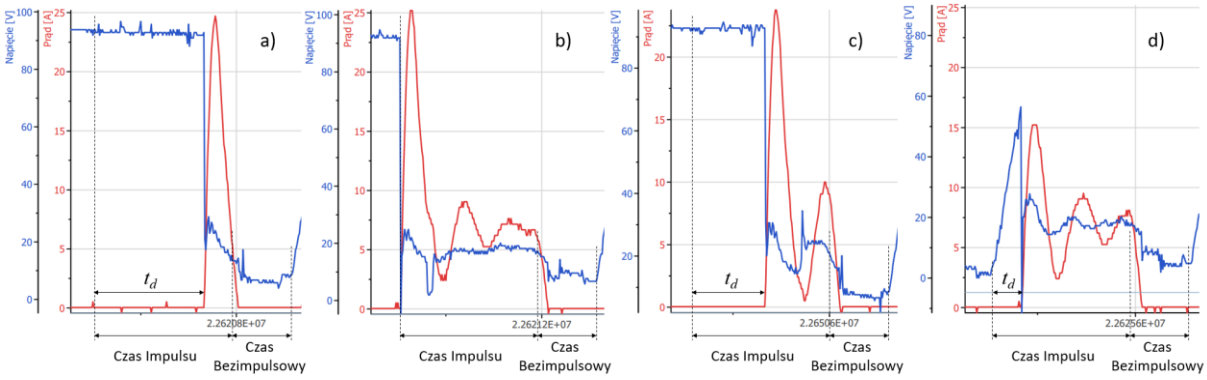
Rysunek 36 pokazuje przykładowy wycinek przebiegu drążenia z okresu ok. 7ms dla drążenia z parametrami #1. Generacja poszczególnych impulsów realizowana jest cyklicznie z częstotliwością określona przez czasy T_{on} i T_{off} (w przykładzie odpowiednio 70µs i 30µs). Każdy z okresów T_{on}, w teorii pozwala na jonizację dielektryka i generację iskry. Natomiast ze względu na dużą zmienność czynników procesu, takich jak odległość najbliższych punktów elektrody i detalu w czasie T_{on}, ilości zanieczyszczeń w szczelinie EDM, stan dielektryka w szczelinie czy możliwość wystąpienia styku elektrody i obiektu, w każdym okresie T_{on} istnieje pewne zmienne prawdopodobieństwo zaistnienia każdego z typów impulsów.

Tak duża zmienność procesu utrudnia jednoznaczną interpretację stanu procesu na podstawie pojedynczych impulsów, czy nawet na podstawie krótkich wycinków danych.

Dopiero dłuższa analiza trendów zmienności częstotliwości występowania każdego z wyładowań o określonym charakterze, może dawać prawdziwą informację o etapie drążenia.

Dokonując przeglądu samych wyładowań, zaobserwowano różne typy impulsów normalnych z pewnymi różnicami w porównaniu do wzorców opisywanych w literaturze [49-52]. Poniżej (Rysunek 37) przedstawiono przykładowe przebiegi impulsów normalnych.

Impulsy normalne także wyróżniały się dużą zmiennością. Co do zasady, większość z impulsów generowana była w czasie impulsu T_{on} po czym była wygaszana w czasie T_{off}, aby umożliwić płukanie szczeliny.



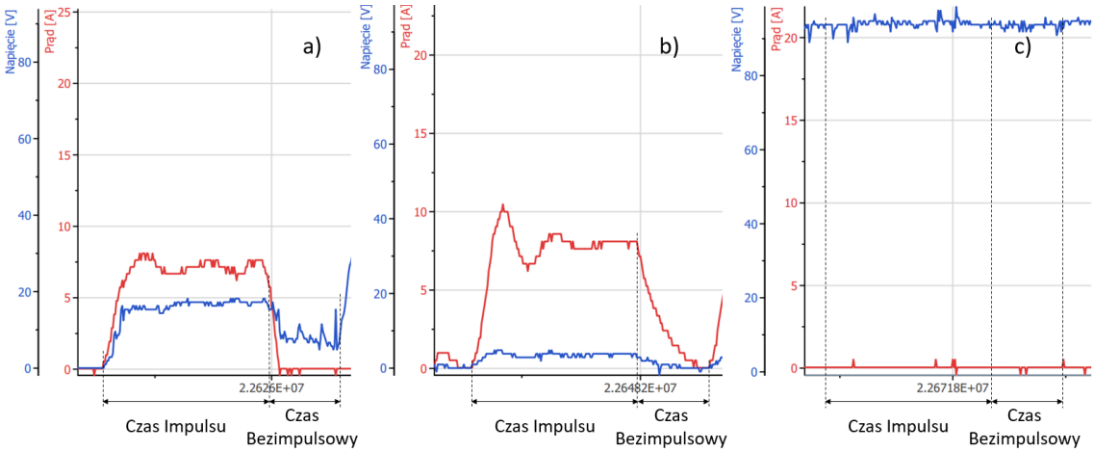
Rysunek 37 Przykładowe zarejestrowane typy impulsów normalnych (zestaw parametrów #1)

Duża część impulsów charakteryzowała się występowaniem pewnego czasu opóźnienia impulsu t_d, od rozpoczęcia czasu impulsu T_{on}. Czas ten był różny w wypadku różnych impulsów. Pomimo umożliwienia generacji impulsu, czasami dielektryk potrzebował dużego czasu na jonizację. Jeśli jonizacja po czasie t_d, zachodziła przy jednoczesnym maksymalnym napięciu szczeliny, zwykle obserwowany był impuls prądowy o dużej wartości (ok 25A) połączony jednocześnie ze spadkiem napięcia (Rysunek 37 a). W części impulsów przełamanie dielektryczne następowało praktycznie zaraz na początku czasu T_{on} (Rysunek 37 b). Sytuacja ta miała miejsce zwykle, kiedy początek czasu impulsu zachodził przy wartości napięcia bliskiej maksymalnej. W tym wypadku, obserwowano duży impuls prądowy połączony z utrzymaniem przepływu prądu na mniej więcej ustalonym poziomie w reszcie czasu T_{on}. Rysunek 37 c) pokazuje przykład sytuacji pośredniej, gdzie przełamanie dielektryczne nastąpiło w połowie okresu T_{on} oraz występował wtórny impuls o małej amplitudzie. W zarejestrowanych impulsach normalnych, zakończenie czasu impulsu i przejście do czasu bezimpulsowego T_{off} owocowało szybkim zanikiem przepływu prądu i zmniejszeniem napięcia do małej wartości w okolicach 10V. Czas bezimpulsowy powinien pozwalać na efektywne płukanie szczeliny i usuwanie

zanieczyszczeń powstałych przy generacji iskry. W okresach następujących bezpośrednio po wyładowaniach obserwowano powolny przyrost wartości napięcia z powrotem do wartości maksymalnej. Czasami, jeśli warunki procesu takie jak stan dielektryka w szczelinie, ilość zanieczyszczeń czy odległość elektrody od obiektu na to pozwalały, następowało generowanie kolejnego impulsu normalnego lub wystąpienia impulsu łukowego bądź zwarcia jeszcze przed osiągnięciem wartości maksymalnej napięcia (Rysunek 37 d). W tym wypadku wartość szczytowa prądu była proporcjonalna do aktualnej wartości napięcia. Pokazane na rysunku przebiegi są tylko przykładami najbardziej typowych impulsów normalnych. W każdym z zarejestrowanych przebiegów próbnych zaobserwować można pewną zmienność czasów opóźnienia t_d oraz wartości napięcia na początku generacji impulsu, a tym samym wartości maksymalnej impulsu prądowego.

W zarejestrowanych wierceniach testowych praktycznie wszystkie impulsy normalne były generowane w wyniku przełamania dielektrycznego. Nie obserwowano drugiego typu impulsów normalnych często opisywanych w literaturze [52] tzn. impulsów generowanych w wyniku przełamania termicznego. Co prawda w części impulsów obserwowano typowy dla tego typu wyładowań opóźniony czas wystąpienia, natomiast nie występował poprzedzający go przepływ niedużego prądu wstępnego. Brak przepływu prądu wstępnego nie generował efektu Joule'a podgrzewającego dielektryk i nie pozwalał na występowanie impulsów termicznych. Jest to spowodowane najprawdopodobniej intensywnym płukaniem i relatywnie małą powierzchnią szczeliny, w której generowane są wyładowania w procesie FHD. Czynniki te nie pozwalały na miejscowe znaczne podgrzewanie obiektu obrabianego i otaczającego go dielektryka.

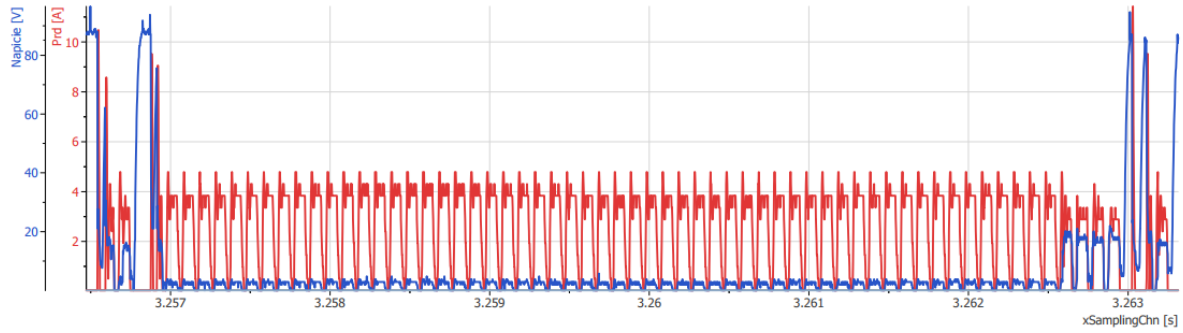
Oprócz impulsów normalnych, w zarejestrowanych danych zidentyfikowane zostały także stany anormalne. Na poniższym rysunku zaprezentowano przykładowe obserwowane przebiegi impulsów: wyładowania łukowego (Rysunek 38 a), zwarcia (Rysunek 38 b) oraz obwodu otwartego (Rysunek 38 c).



Rysunek 38 Zaobserwowane typy impulsów: wyładowanie łukowe (a), zwarcie (b) oraz obwód otwarty (c)

Wyładowanie łukowe (Rysunek 38 a)), charakteryzuje się, zgodnie z teorią, brakiem czasu opóźnienia wyładowania. Już na starcie czasu impulsu T_{on} występuje natychmiastowy wzrost wartości prądu oraz napięcia i utrzymanie ich na stałym poziomie przez cały okres Impulsu. Oznacza to wytworzenie w szczelinie łuku elektrycznego, który jest utrzymywany przez całość okresu T_{on} i zanika w T_{off}. Po zakończeniu czasu impulsu i przejściu do czasu bezimpulsowego T_{off}, następuje wygaszenie łuku połączone ze spadkiem prądu do zera. Wyładowania łukowe powstają ze względu na stan dielektryka w szczelinie, który nie odzyskał jeszcze własności dielektrycznych po poprzednim wyładowaniu lub ze względu na dużą koncentrację zanieczyszczeń. Powodowane jest to zwykle pogorszonymi warunkami płukania.

Przy impulsie zwarciowym (Rysunek 38 b)) zachodzi podobny wzrost i utrzymanie wartości prądu jak przy wyładowaniu łukowym. Zasadniczą różnicą jest towarzyszący zwarciu niski poziom napięcia. W czasie bezimpulsowym podczas zwarcia obserwujemy powolny spadek wartości prądu, zamiast szybkiej zmiany jak przy wyładowaniu łukowym. Z tego względu monitorowanie zarówno wartości prądu jak i napięcia może ułatwiać rozróżnianie tych dwóch stanów. Zwarcia są wynikiem zetknięcia powierzchni elektrody z detalem obrabianym. W tym czasie nie zachodzi generowanie normalnych iskier przez co drążenie nie jest prawidłowo kontynuowane (materiał elektrody i detalu nie jest erodowany). Ze względu na to, w przebiegach obserwowano występowanie zwarć w seriach trwających zwykle od 3ms do 16ms. Przykład takiej serii impulsów zwarciowych przedstawiono na poniższym rysunku 39.



Rysunek 39 Seria wyładowań zwarciowych

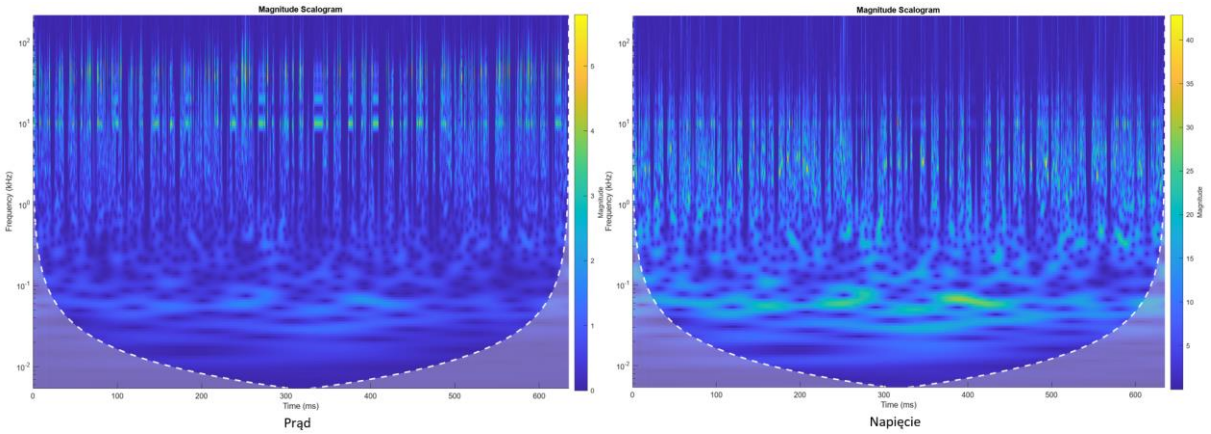
W celu przerwania serii wyładowań zwarciowych konieczne jest wycofanie elektrody wzdłuż osi Z i wznowienie drążenia z nowej pozycji, bez styku. Czynność ta realizowana jest automatycznie przez serwomechanizm EDM maszyny.

Stan obwodu otwartego (Rysunek 38 c)) charakteryzuje się utrzymaniem maksymalnej lub rosnącej wartości napięcia i zerowej wartości prądu przez cały cykl T_{on}, czyli całkowity brak generacji iskry w szczelinie EDM. Obserwowane drążenia posiadały wiele okresów EDM, w których nie zachodziło generowanie żadnego z typów wyładowań. Stany te utrzymywały się przez różne czasy, od pojedynczych okresów EDM do długich odcinków bezimpulsowych osiągających nawet ponad 100 kolejnych okresów. Sytuacja ta wynikała najprawdopodobniej z dużego obciążenia cyklu dla ustawień procesu. W tych warunkach nawet relatywnie rzadkie występowanie wyładowań może nadal dawać dobrą szybkość obróbki. Dodatkowo warunki takie mogą dawać dobre warunki płukania, co poprawia własności dielektryka i zapobiega łatwemu tworzeniu łuków i zwarć. W tych sytuacjach serwomechanizm EDM musi pomniejszać szczelinę, aby wznowić generację iskier.

Zasadniczo, w żadnym z wierceń testowych nie zaobserwowano tzw. impulsów obwodu otwartego [52], które charakteryzować powinny się napięciem bliskim maksymalnemu, ale dodatkowo przepływem w obwodzie małego prądu o stałej amplitudzie. Oznacza to, że w obserwowanych drążeniach mieliśmy do czynienia z czystym procesem EDM. Nie obserwowano warunków procesu, które pozwalałyby na elektrochemiczne erodowanie materiału (proces ECM).

Analiza danych zebranych z wiercenia otworów z zastosowaniem parametrów agresywnych (parametry #2) oraz różnych wariantów wierceń (wiercenia pod kątem itd.), wykazała podobny charakter zachodzących wyładowań. Zasadniczą różnicą pomiędzy

wariantami procesów była częstotliwość występowania poszczególnych typów wyładowań. Sama zmiana częstości występowania impulsów w istotny sposób wpływała na globalne efekty drążenia jak jakość powierzchni, czy czas drążenia pojedynczego otworu.



Rysunek 40 Skalogram przebiegu prądu i napięcia dla przykładowego fragmentu drążenia otworu testowego

W celu dogłębniejszego poznania charakteru zarejestrowanych przebiegów, na ich fragmentach danych dokonano analizy metodą ciągłej transformaty falkowej CWT (ang. Continous Vawelet Transform). Na rysunku 40 pokazano przykładowe skalogramy dla przebiegów napięcia i prądu. Zobrazowano 650ms z etapu III drążenia (normalne drążenie wewnątrz płytki 1).

Na przebiegu prądu zaobserwować można obszary o zwiększonej amplitudzie i częstotliwości w okolicy 10kHz. Potwierdza to korelację generacji iskier EDM z cyklicznością okresów T_{on} i T_{off}, ponieważ w obrazowanym drążeniu okres T wynosił 100µsm (T_{on} i T_{off} odpowiednio 70µs i 30µs), co daje częstotliwość fali nośnej generatora EDM na poziomie 10kHz. Obserwowalne są także pewne zagęszczenia składowych o wyższych częstotliwościach co może wynikać z dużej zmienności kształtu samych impulsów. Na przebiegu napięcia widać dodatkowo dużą zawartość składowych o niskich częstotliwościach ok. 10^{-1}Hz. Jest to skorelowane najprawdopodobniej z dużą obserwowaną zawartością obwodów otwartych podczas drążenia. W tym stanie napięcie często utrzymywało dużą wartość (często dochodzącą do maksimum), natomiast nie występował wtedy przepływ prądu, przez co na skalogramie prądu składowe o niskich częstotliwościach mają małe amplitudy. Analizy skalogramów potwierdzają wcześniejsze wnioski o dużej zmienności i stochastycznym charakterze generacji iskier w czasie procesu wyciągnięte na podstawie przeglądu przebiegów czasowych.

4.6. Synchronizacja danych z różnych systemów

Jednym z problemów napotkanych w analizie, była konieczność synchronizacji zarejestrowanych danych wysokoczęstotliwościowych przebiegów napięcia i prądu oraz danych rejestrowanych przez sterownik elektrodrążarki. W czasie prowadzenia drążeń testowych systemy nie były z sobą synchronizowane, rejestracja zaczynała i kończyła się w różnych momentach procesu oraz zapisane sygnały miały różne próbkowania. Z tych względów konieczne było zastosowanie metody pozwalającej na stworzenie zsynchronizowanych czasowo plików wyjściowych dających możliwość analizy zależności różnych mierzonych wartości.

W tym celu zastosowano metodę korelacji rejestrowanych przez oba systemy przebiegów napięcia jako jedynej wartości zapisywanej przez oba systemy.

W celu korelacji przeprowadzono proces wygładzania napięcia wysokoczęstotliwościowego zgodnie z metodologią opisaną w rozdziale 4.6 (Równanie 44). Każdą z wartości sygnału zastąpiono nową wartości stworzoną z uśrednienia 25 000 sąsiadujących próbek. Podobnie napięcie szczeliny rejestrowane przez maszynę wygładzono uśredniając 8 sąsiednich próbek. Osiągnięto w ten sposób 2 sygnały z różnych systemów przedstawiające odpowiadające sobie te same wartości. Niestety, powstałe sygnały posiadały nadal różne częstotliwości próbkowania. Aby porównywać oba przebiegi dokonano procesu zmniejszenia częstotliwości próbkowania (downsamplingu) obu sygnałów. Każdy z przebiegów został wstępnie poddany interpolacji metodą krzywych sklejanych Akima [310]. Interpolacja Akimy zbudowana jest z kawałkowych wielomianów trzeciego stopnia. Do określenia współczynników wielomianu interpolacyjnego używane są tylko dane z sąsiednich punktów. Nie ma potrzeby rozwiązywania dużych układów równań, dlatego ta metoda interpolacji jest bardzo wydajna obliczeniowo. Ponieważ nie zakłada się funkcjonalnej postaci dla całej krzywej, a uwzględnia jedynie niewielką liczbę punktów, metoda ta nie prowadzi do sztucznych oscylacji w wynikowej krzywej. Monotoniczność określonych punktów danych nie jest koniecznie zachowywana przez uzyskaną funkcję interpolacyjną. Przez dodatkowe ograniczenia na oszacowane pochodne można skonstruować funkcję interpolacyjną zachowującą monotoniczność [311].

Dla danego zestawu punktów:

$$s_i = s(x_i), \qquad 1 \leq i \leq k, \tag{16}$$

Funkcja interpolacyjna jest definiowana jako:

$$s(x) = a_0 + a_1 \cdot (x - x_i) + a_2 \cdot (x - x_i)^2 + a_3 \cdot (x - x_i)^3, \qquad x_i \leq x \leq x_{i+1} \tag{17}$$

Aby określić współczynniki a_0, a_1, a_2 i a_3 wielomianu interpolacyjnego dla każdego przedziału $[x_i, x_{i+1}]$, używane są wartości funkcji s_i i s_{i+1} oraz pierwsze pochodne $s_i{}'$ i $s_{i+1}{}'$ na końcach przedziału.

Pierwsza pochodna $s_i{}'$ funkcji interpolacyjnej w punkcie x_i jest oszacowana na podstawie danych dla tego punktu i dwóch następnych punktów po obu stronach x_i. Korzystając z proporcji:

$$d_j = \frac{s_{j+1} - s_j}{x_{j+1} - x_j}, \qquad j = i - 2, i - 1, i, i + 1 \tag{18}$$

oraz współczynników wagowych:

$$w_{i-1} = |d_{i+1} - d_i|, \tag{19}$$

$$w_i = |d_{i-1} - d_{i-2}|, \tag{20}$$

oszacowana pochodna $s_i{}'$ jest definiowana jako:

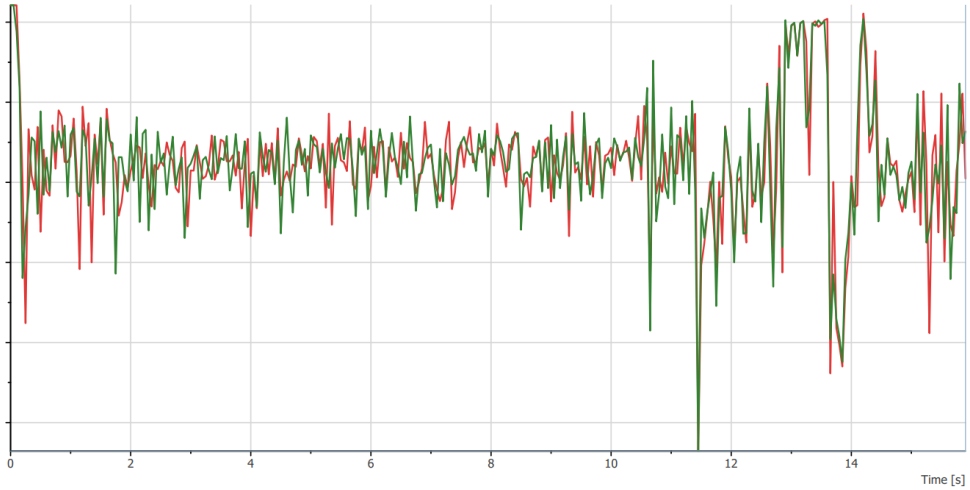
$$s_i' = \frac{w_{i-1} \cdot d_{i-1} + w_i \cdot d_i}{w_{i-1} + w_i} \tag{21}$$

Tak zbudowane krzywe interpolacyjne z obu przebiegów zostały poddane ponownemu próbkowaniu z częstotliwością 20 S/s. Ponieważ wartości napięcia rejestrowane przez maszynę były zapisywane do plików bez skalowania do jednostek inżynierskich przeprowadzono normalizację obu przebiegów napięcia do wartości (0,1). Kolejnym krokiem było zgranie czasowe tak stworzonych sygnałów. W tym celu zastosowano przesuwanie każdego z przebiegów względem siebie i cykliczne określanie ich korelacji metodą Pearsona.

Współczynnik korelacji Pearsona [312] to miara, określająca siłę liniowego związku między dwiema zmiennymi. Oblicza się go przez podzielenie kowariancji zmiennych x i y przez iloczyn odchylenia standardowego każdej zmiennej, zgodnie ze wzorem:

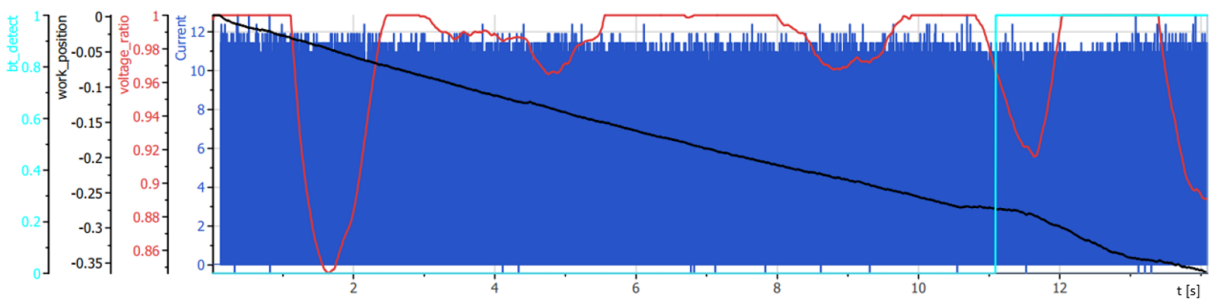
$$r = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2 \sum(y - \bar{y})^2}} \tag{22}$$

Przebiegi były ustawiane względem siebie w położeniu wykazującym największą korelację między sygnałami. Na tej podstawie określono wzajemne położenie wszystkich sygnałów danych. Przykład skorelowanych danych pokazano na rysunku 41.



Rysunek 41 Skorelowane resamplowane przebiegi napięć maszyny i systemu pomiarowego

Próbki nadmiarowe wykraczające poza zakres skorelowanych danych w pliku wysokoczęstotliwościowym zostały usunięte z plików wynikowych. W ten sposób zbudowano zbiorcze, skorelowane czasowo pliki wszystkich zarejestrowanych danych dla każdego wierconego otworu testowego. Przykładowe przebiegi skorelowanych czasowo danych dla otworu R5 H11 przedstawiono na rysunku 42.



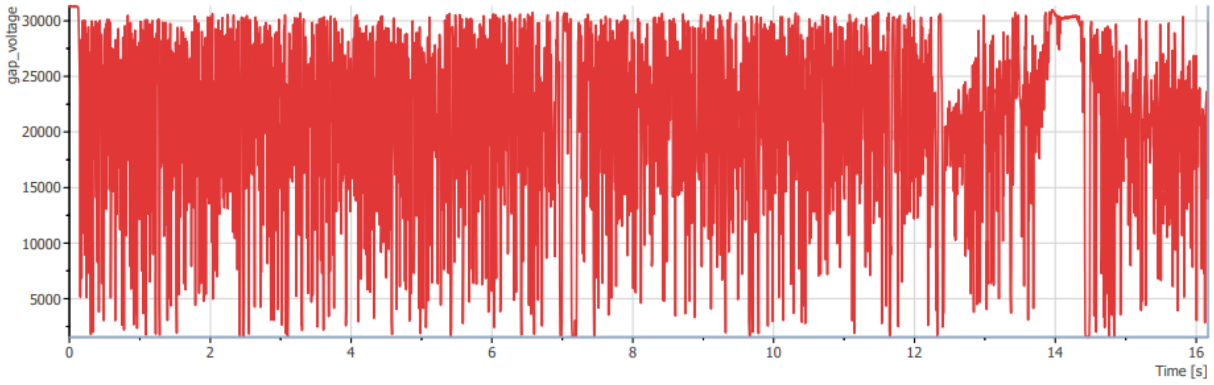
Rysunek 42 Przykładowe przebiegi skorelowanych czasowo danych. Kolor czerwony - zmiana napięcia, niebieski - napięcie wysokoczęstotliwościowe, czarny - przemieszczenie wrzeciona, jasnoniebieski – szacowane wykrycie przebicia

Tak stworzone pliki posłużyły do dalszej analizy danych i opracowania metod wykrywania przebicia.

4.7. Analiza przebiegów rejestrowanych przez sterownik maszyny

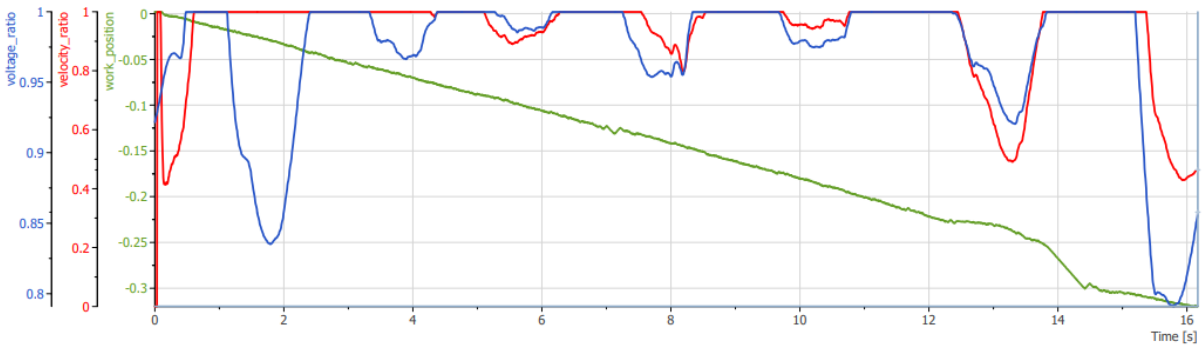
Mając zsynchronizowane dane z maszyny można dokonać analizy ich przydatności w procesie monitorowania etapu procesu. Jak opisywano w rozdziale 3.1, część z badaczy [62, 63] opierało swoje podejścia o analizę niskoczęstotliwościowych parametrów opisujących warunki procesu. Dokonano przeglądu danych rejestrowanych przez sterownik maszyny MKG 1000.

Jednym z potencjalnie przydatnych sygnałów może być napięcie procesu. Maszyna rejestruje nieprzetworzone napięcie szczeliny, natomiast ze względu na niską częstotliwość zapisu do pliku wyjściowego nie daje prawdziwych informacji o procesie. Przykładowy przebieg napięcia pokazano na rysunku 43. Sygnał charakteryzuje się bardzo dużą zmiennością. Najprawdopodobniej ze względu na zjawisko aliasingu wartość napięcia zapisana do pliku przyjmuje w dużej mierze losowe wartości. Tak zapisana wartość napięcia szczeliny nie dostarcza więcej informacji niż wygładzone wartości napięcia opisywane wcześniej.



Rysunek 43 Zarejestrowane przez maszynę napięcie szczeliny

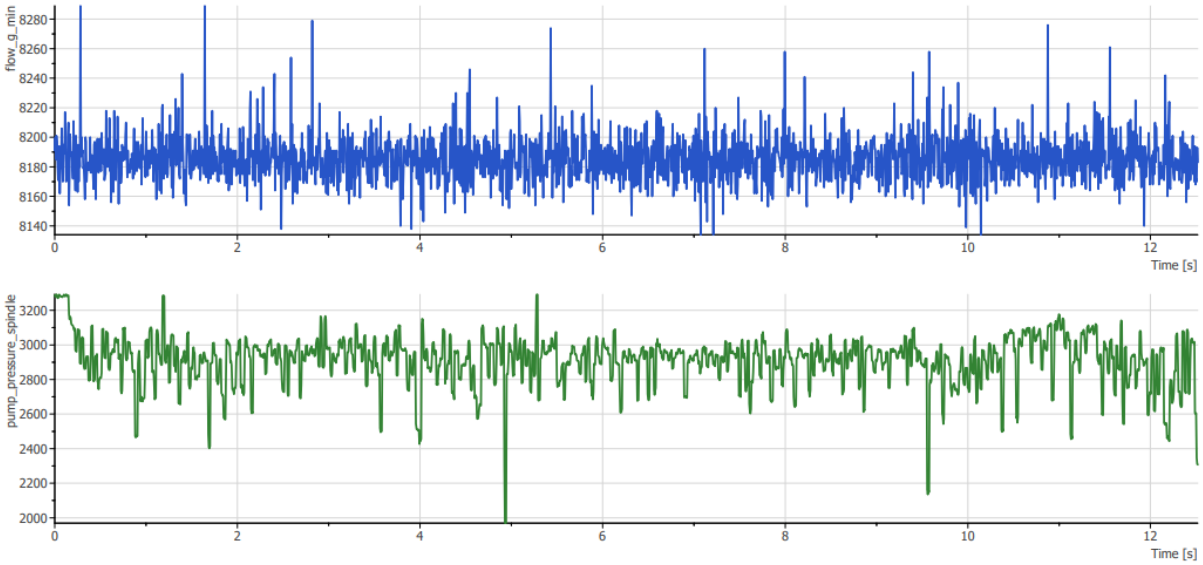
Interesujące z punktu widzenia monitorowania procesu mogą być wartości stosunku napięcia, stosunku przemieszczenia i położenia wrzeciona zapisywane przez sterownik maszyny. Wartości stosunków obrazują tempo zmiany danej wartości. Porównują one uśrednione wartości z poprzedniego okresu do obecnego okresu o określonej długości.



Rysunek 44 Rejestrowane przez maszynę przebiegi stosunku napięcia, stosunku przemieszczenia i pozycję względną wrzeciona.

Część z badań [64, 65, 69] opierała swoje metody na założeniu o zmienności charakteru wyładowań w momencie przebicia. Dowodzili oni, że w tym etapie procesu ilość obwodów otwartych powinna się zwiększać. Zjawisko takie powinno być widoczne także w przebiegach różnic sygnałów zapisanych przez maszynę. Na zapisanych przebiegach zaobserwować można korelację zmian napięcie i zmian pozycji wrzeciona w okolicach punktu przebicia otworu (Rysunek 44 - około 13s drążenia na wykresie). Natomiast jak opisywano wcześniej, wartości te wymagają uśredniania i porównywania ich zmiany w dłuższym okresie, co może powodować opóźnienia wykrywania przebicia. Dodatkowo same warunki prowadzenia procesu mogą spowodować generowania impulsów o charakterze wpływającym na obniżenie średniej wartości napięcia (np. poprzez zwiększenie ilości obwodów otwartych, lub dużą ilość zwarć) co może prowadzić do błędnego wykrycia przebicia. W czasie prawidłowo prowadzonego procesu wartości stosunków zmieniają się jak widać na powyższym przykładzie.

Analiza sygnału pozycji wrzeciona, powinna być obarczona mniejszym opóźnieniem (brak uśredniania i porównywania wartości). Wartość przemieszczenia jest związana bezpośrednio z wartością napięcia szczeliny jak opisano w rozdziale 2.3.1. Podobnie jak z wartością stosunku napięcia, różne konfiguracje impulsów mogą prowadzić do błędnych wniosków. Dodatkowo zużycie elektrody i erozja materiału z detalu obrabianego mogą następować z różną prędkością pomiędzy drążeniami, dlatego opieranie wykrywania przebicia jedynie na wartości przemieszczenia wrzeciona nie da wystarczająco poprawnych wyników.



Rysunek 45 Wartości przepływy i ciśnienia dielektryka pompowanego przez elektrodę w czasie drążenia

Część z badaczy [62] próbowało opierać swoje analizy na interpretowaniu zmian w sygnale ciśnienia dielektryka tłoczonego przez elektrodę. Jak pokazano na rysunku 45, w wielu wypadkach rejestrowane wartości przepływu i ciśnienia dielektryka pompowanego przez wnętrze elektrody w czasie drążenia nie dostarcza jednoznacznych informacji o etapie procesu (brak wyraźnej zmiany wartości po przebiciu). W części z analizowanych danych można zaobserwować niewielką zmianę wartości ciśnienia, najprawdopodobniej skorelowaną z momentem wystąpienia przełomu. Niestety ze względu na dużą zmienność wartości, brak powtarzalności w różnych drążeniach i małą częstotliwość próbkowania ciężko wyciągać jednoznaczne wnioski na podstawie tych danych.

Przegląd wszystkich przebiegów testowych potwierdza hipotezę, że analiza samych danych niskoczęstotliwościowych rejestrowanych przez maszynę nie dostarcza dostatecznych informacji do identyfikacji etapu procesu i wykrywania przebicia.

4.8. Podsumowanie analizy drążeń testowych FHD

W pracy dokonano analizy przebiegów prądu i napięcia rejestrowanych w czasie drążenia FHD otworów wentylacyjnych. Wykonano analizę przebiegów w skali makro (całego procesu) i mikro (pojedynczych wyładowań). Przegląd charakteru krótkich odcinków danych pozwolił na zidentyfikowanie typowych impulsów występujących w czasie procesu. Potwierdzona została hipoteza o zasadniczym podobieństwie tych wyładowań do impulsów opisywanych

w literaturze [49-52] dotyczącej także innych procesów EDM (sinker i WEDM). W rozprawie opisano dokładniej cechy impulsów znalezionych w przebiegach testowych FHD.

Opisana analiza przeprowadzonych drążeń testowych obrazuje złożoność zagadnienia prawidłowego wykrywania przełomu i całkowitego przebicia otworu w procesie FHD. Bardzo duża zmienność procesu generacji wyładowań efektywnie utrudnia prawidłowe interpretowanie wyników tylko na podstawie sztywno ustalonych wartości progowych rejestrowanych wielkości lub na podstawie uśrednianych lub filtrowanych przebiegów czasowych. W związku z tym konieczna jest bardziej dogłębna analiza charakteru poszczególnych impulsów występujących w różnych etapach procesu oraz opracowanie metod wykrywania przebicia biorących tą charakterystykę pod uwagę. Opis tych zagadnień przedstawiono w następnym rozdziale.

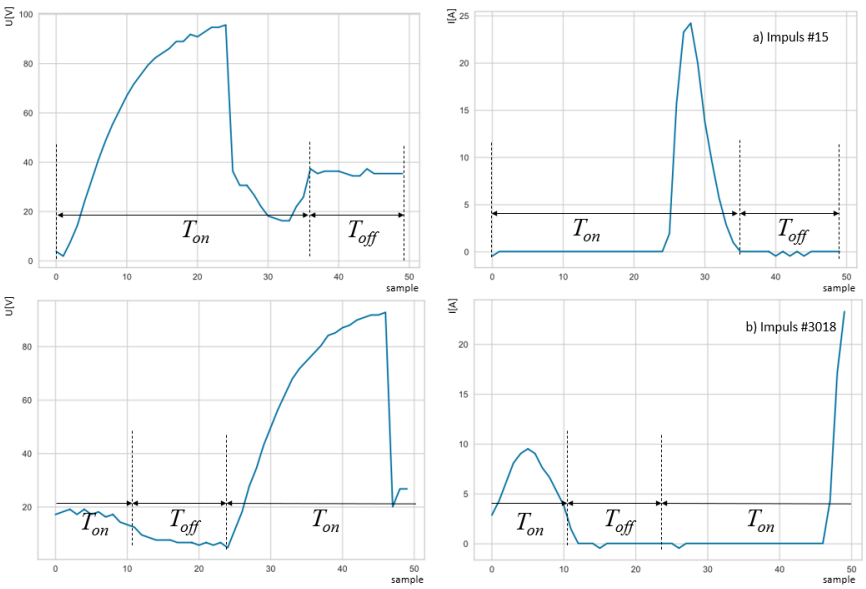
5. Opracowanie narzędzi rozpoznawania przebicia i anomalii procesu

5.1. Metodologia zautomatyzowanej analizy impulsów procesu FHD

Aby lepiej zrozumieć charakter procesu drążenia podjęta została próba analizy charakteru przebiegów prądu i napięcia mająca na celu określenia jakie typy impulsów występują w różnych etapach drążenia. Opracowanie narzędzia klasyfikującego impulsy może dodatkowo posłużyć jako narzędzie diagnostyczne w procesie analizowania przyczyn powstawania defektów lub podczas testowania nowych zestawów parametrów drążenia. W tym celu zarejestrowane przebiegi napięcia oraz prądu poddane zostały obróbce mającej na celu wyodrębnienie poszczególnych impulsów prądowych.

5.1.1. Podział zestawów danych na impulsy

Jak opisano w rozdziale 4.5 poszczególne impulsy wykazują dużą zmienność, natomiast powinny zachodzić z określoną, odgórnie określoną częstotliwością wyznaczoną przez okres impulsów *T*, złożony z czasów T_{on} *i* T_{off}. Oznacza to, że można podzielić czasowe przebiegi prądu i napięcia na mniejsze zestawy danych jednoznacznie odpowiadające poszczególnym impulsom. Analizując przebiegi można zidentyfikować w nich impulsy z wyraźnie określonymi granicami T_{on} i T_{off}. Na podstawie nich możliwy był podział całości plików poprzez podział ich na pakiety odpowiadające okresowi T (odpowiednio 100µs i 60µs dla dwóch zestawów parametrów). Niestety takie podejście nie przyniosło oczekiwanych efektów. Prosty podział pliku na kolejne sekwencyjne zestawy o czasie trwania T, powodował obserwowalne przesuwanie granic impulsów. Każdy z kolejnych zestawów danych wykazywał rosnące przesunięcie czasu rozpoczęcia i zakończenia impulsu. W plikach zawierających tak dużą ilość impulsów efektywnie uniemożliwiało to analizę impulsów. Zaobserwowano, że już po ok. 6900-7100 okresów (czyli po 700ms dla parametrów #1) przesunięcie czasowe odpowiadało jednemu całemu okresowi impulsu. Poniższy rysunek 46 przedstawia przykładowe przesunięcie czasowe przy podziale impulsów na paczki co 100µs (drążenie R5 H7).



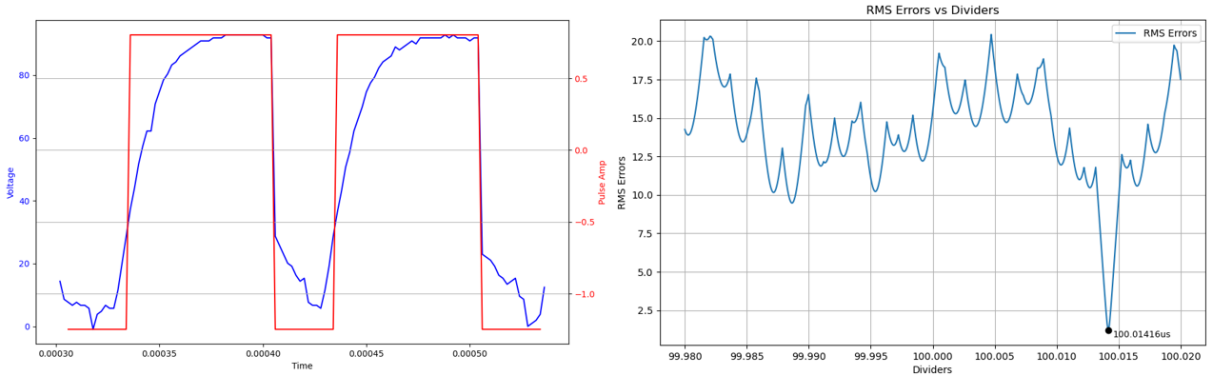
Rysunek 46 Przesunięcie czasowe impulsów a) impuls #15 b) impuls #3018

Przesunięcie czasowe wynikało z braku synchronizacji podstawy czasu maszyny oraz układu pomiarowego (kontrolera PXI). Najprawdopodobniej, oscylatory kwarcowe maszyny oraz sterownika charakteryzują się różnymi częstotliwościami taktowania, co w efekcie wprowadza różnice czasowe. Okres 100 µs generowany przez maszynę jest minimalnie większy niż teoretyczny okres 100 µs zarejestrowany przez sterownik PXI. Aby uniknąć tego problemu po raz kolejny wykorzystano metodę korelacji Personsa (rozdział 4.6). Stworzono dwa wzorcowe impulsy napięcia odpowiadające dwóm następującym po sobie okresom wzrostu napięcia (użyto wyidealizowanego przebiegu prostokątnego jako wzorca). Następnie zbudowano algorytm wyszukujący w całym pliku, fragmentów przebiegów o najlepszej korelacji do przebiegów wzorcowych. W całych przebiegach zarejestrowanych, okno odpowiadające dwóm okresom EDM było korelowane z impulsem wzorcowym. Okno było cyklicznie przesuwane o jedną próbkę i korelacja była powtarzana do momentu przeskanowania całego pliku. Algorytm zapamiętywał numery próbek dla początków skanowanych okien o największym współczynniku korelacji do impulsu wzorcowego. Zapamiętywano impulsy wykazujące korelację lepszą niż 95% całej populacji. Znając numery próbek początków impulsów najlepiej odpowiadających wzorcowi, można było określić przesunięcie czasowe pomiędzy zegarami maszyny i sterownika. Określono odległości pomiędzy znalezionymi fragmentami sygnału i ustalono najlepszy wspólny dzielnik z zakresu ± 2% od teoretycznej ilości próbek przypadającej na daną częstotliwość próbkowania. Algorytm najpierw tworzył tablicę 100001 wartości wszystkich możliwych dzielników z zakresu.

Dla każdego z nich określano błąd RMS wg wzoru:

$$e_{rms} = \sqrt{(c_i/d_j)^2} \tag{23}$$

Gdzie c_i to odległość między dobrze skorelowanymi impulsami a d_j to wartość dzielnika z badanego zakresu. Dzielnik o najmniejszej wartości błędu RMS był zapamiętywany dla potrzeb korekcji przesunięcia zegarów maszyn. Znaleziony optymalny dzielnik odpowiadał realnej ilości próbek rejestrowanych przez system PXI przypadających na jeden okres EDM.



Rysunek 47 Wykres przebiegu napięcia zarejestrowanego o najlepszej korelacji do przebiegu wzorcowego (po lewej) oraz wykres błędu RMS dla dzielników (po prawej)

Analizę powtórzono dla wielu plików wejściowych. Na potrzeby analizy pliki zostały resamplowane do tej samej częstotliwości. Każdy z analizowanych plików dawał zbliżoną wartość różnicy czasowej. Poniżej stablicowano wartości dzielników dla każdego z pliku.

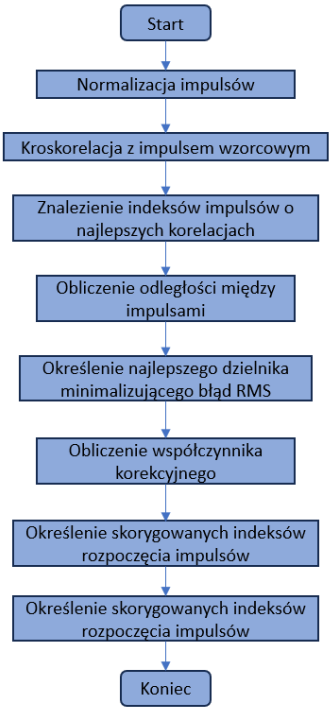
Tabela 7 Najlepsze dzielniki (ilość zarejestrowanych próbek na okres EDM)

	Row1	Row2	Row3	Row4
Hole 1	50,0070422	50,0070426	50,0071346	50,0070822
Hole 2	50,0070366	50,0070286	50,00713	50,0070342
Hole 3	50,0070604	50,0070758	50,0070688	50,007046
Hole 4	50,0070752	50,0070194	50,0071248	50,0070914
Hole 5	50,007026	50,0070222	50,0070184	50,0070612
Hole 6	50,0069914	50,0070152	50,0070688	50,0070278
Hole 7	50,0070198	50,0070442	50,007015	50,0071376
Hole 8	50,0071026	50,0071026	50,007105	50,0070864
Hole 9	50,0070334	50,0071136	50,0070756	50,0070632
Hole 10	50,007002	50,0070592	50,0070526	50,007021

Jak widać, najlepsze dzielniki wahają się w okolicach średniej wartości 50,00705894. Przekłada się to na przesunięcie zegarów w okolicach 0,705894µs na próbkę (przy częstotliwości próbkowanie 500 kS/s). Potwierdziło to wstępne obserwacje przesunięć

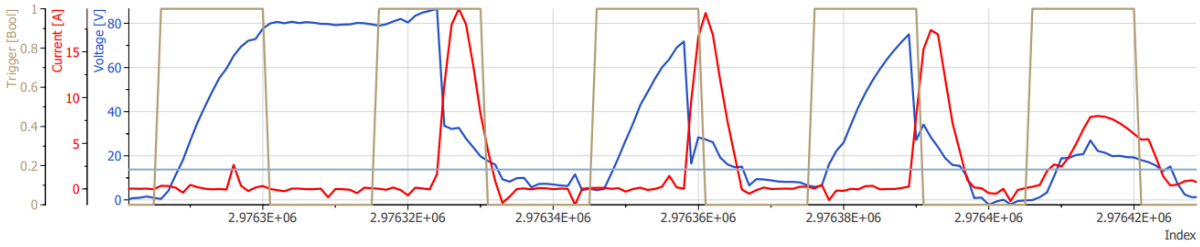
czasowych przy podziale sygnałów na impulsy. Wg obliczeń po zarejestrowaniu ok. 7083 próbek przesunięcie czasowe będzie odpowiadało jednemu okresowi EDM.

Znając wartość różnicy czasowej urządzeń można było zaimplementować tą wiedzę do podziału pliku na zestawy danych. Wdrożono algorytm kompensujący przesunięcia. Po podziale określonej liczby zestawów danych, odpowiadającej kumulacji przesunięcia czasowego do wartości odpowiadającej przesunięciu o jedną próbkę, wprowadzano przesunięcie indeksu próbki rozpoczęcia kolejnego impulsu o jeden indeks do przodu. W rezultacie uzyskano zestawy danych dobrze odpowiadające okresom impulsów.



Rysunek 48 Algorytm synchronizacji czasów impulsów

Wartość przesunięcia czasowego została potwierdzona w późniejszych pomiarach z zastosowaniem karty pomiarowej PXIe-6363. Do pomiarów prądu i napięcia dodano pomiar przebiegu sygnału prostokątnego generowanego przez maszynę i podawanego do układu tranzystorowego odpowiedzialnego za generowanie okresów T_{on} i T_{off}. Sygnał ten pozwolił na obiektywne określenie okresu EDM generowanego przez maszynę.



Rysunek 49 Przebieg prądu, napięcia i sygnału generującego Ton i Toff

Pomiary potwierdziły wartości przesunięć czasowych określonych opisaną wyżej metodą analityczną. Prostokątny sygnał generatora EDM rejestrowany układem pomiarowym 2 służył także do podziału plików na impulsy w późniejszych badaniach. Ponieważ sygnał rejestrowany był przez ten sam system, który rejestrował też napięcie i prąd, różnice częstotliwości zegarów maszyn nie wprowadzały błędu przesunięcia czasowego. Wykrywanie zbocz rosnących i opadających jednoznacznie definiowało granice czasów T_{on} i T_{off}.

5.1.2. Podział impulsów na kategorie (klastrowanie impulsów)

Posiadając tak przygotowane zestawy danych można podjąć próbę podziału całego zestawu danych na podgrupy odpowiadające poszczególnym typom wyładowań. Ze względu na ogromną ilość danych, konieczne było zastosowanie metody pozwalającej na automatyzację tego procesu. Podział na kategorie impulsów można wykonać np. z pomocą procesu klasyfikacji metodami uczenia maszynowego opisywanymi w rozdziale 3.3. Natomiast, te metody uczenia nadzorowanego wymagają wstępnego przygotowania danych polegającego na nadaniu etykiet odpowiadających kategoriom. Ręczne oznaczanie danych byłoby procesem bardzo pracochłonnym. W związku z tym zaproponowano podejście nienadzorowanego uczenia realizujące to zadanie.

W celu automatyzacji zadania opisywania impulsów zastosowano analizę skupień (klastrowanie), co pozwoliło na grupowanie zestawów danych odpowiadających poszczególnym impulsom i nadawanie im odpowiednich etykiet. W tym celu zastosowano dwie metody grupowania prototypowego: analizę k średnich (k-means - rozdział 0) oraz grupowanie aglomeratywne [313]. Metoda k-średnich jest nienadzorowanym algorytmem podziału zbioru danych na określoną liczbę klastrów na podstawie podobieństwa ich cech, bez użycia przypisanych wcześniej etykiet jak robi to algorytm kNN opisany w rozdziale 0. W procedurze każdy punkt danych jest przypisany do najbliższego centroidu, tworząc k klastrów. Następnie algorytm oblicza nowe centroidy jako średnią wszystkich punktów danych przypisanych do każdego centroidu. Kroki powtarzane, aż centroidy nie będą się znacznie zmieniać lub osiągnięta zostanie maksymalna liczba iteracji. Klastrowanie aglomeracyjne jest metodą hierarchiczną, która rozpoczyna działanie od traktowania każdego punktu danych jako oddzielnego klastra, a następnie łączy najbardziej podobne klastry wraz z postępem iteracji. W każdej iteracji dwa najbliższe klastry są łączone, tworząc większy klaster. Ten proces kontynuuje się, aż wszystkie klastry zostaną połączone w jeden duży klaster lub do momentu spełnienia pewnego warunku stopu, na przykład określonej liczby klastrów lub pewnego poziomu podobieństwa między nimi. W badaniach zastosowano odległość euklidesową punktów jako miarę odległości między pojedynczymi obserwacjami zbioru danych oraz kryterium łączenia Warda, które określa niepodobieństwo zbiorów jako funkcję odległości par obserwacji w tych zbiorach.

W celu poprawy wydajności i zwiększenia dokładności modelu, przeanalizowano metody klastrowania na danych nieprzetworzonych oraz danych poddanych standaryzacji i normalizacji.

Standaryzacja była prowadzona korzystając ze wzoru:

$$z = \frac{x - \mu}{\sigma} \tag{24}$$

Gdzie:

x – zmienna niestandaryzowana,

μ – średnia z populacji

σ – odchylenie standardowe populacji

Normalizacja MinMax danych wykonana została zgodnie z zależnością:

$$z = \frac{x - \min(x)}{[\max(x) - \min(x)]} \tag{25}$$

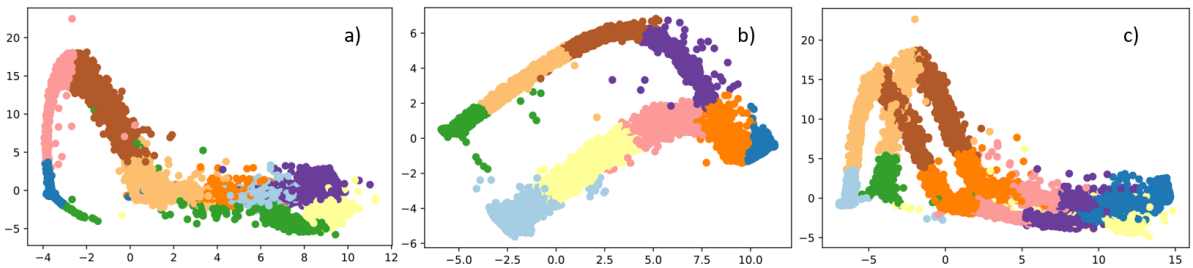
Normalizacja MinMax skaluje dane do przedziału [0,1] przez zmianę danych w populacji tak, aby minimalna wartość wynosiła 0, a następnie dzieliła każdą następną wartość ze zbioru przez maksymalną wartość (czyli różnicę między oryginalną maksymalną i minimalną wartością).

Stosowany algorytm k średnich opiera się na metrykach odległości. Standaryzacja i normalizacja pozwalają zapobiegać temu, aby cechy o większych skalach dominowały w procesie klastrowania.

W metodzie klastorowania k-średnich poprawę wydajności osiągnięto stosując algorytm inicjalizacji k-średnich++ (k-means++) [314] zaproponowany przez D. Arthura et al. Algorytm ten wybiera początkowe centroidy klastrów przy użyciu próbkowania opartego na empirycznym rozkładzie prawdopodobieństwa wkładu punktów w ogólną bezwładność. Ta technika pozwoliła na przyśpieszenie zbieżności obliczeń. W badaniach użyto algorytmu „chciwego k-means++". Różni się on od zwykłego k-means++ przeprowadzeniem kilku prób w każdym kroku próbkowania i wyborem najlepszego centroidu spośród nich. Do liczenia

k-średnich użyto algorytmu „elkan” opracowanego przez C. Elkana [315]. Algorytm Elkan to ulepszona wersja algorytmu k-means, która wprowadza techniki optymalizacyjne, aby zmniejszyć liczbę obliczeń potrzebnych do przypisania punktów do klastrów. W tradycyjnym k-means w każdej iteracji liczone są odległości między tym punktem, a wszystkimi centroidami. Algorytm Elkan wykorzystuje nierówności trójkąta do ograniczenia zbędnych obliczeń, co sprawia, że jest bardziej wydajny obliczeniowo. Dzięki temu znacznie przyspieszył proces klastrowania, szczególnie w wypadku badanych dużych zbirów danych. Co prawda algorytm ten rezerwuje większą ilość pamięci ze względu na alokację dodatkowej tablicy o kształcie ilość_próbek x ilość_klastrów, natomiast w wypadku klastrowania na relatywnie niedużą ilość grup nie stanowiło to problemu podczas realizacji badań.

Klastrowaniu poddane zostały trzy zestawy danych: przebiegi napięcia, przebiegi prądu oraz przebiegi sklejonych wartości prądu i napięcia. Zastosowanie klastrowania bezpośrednio na standaryzowanych danych nie pozwoliło na prawidłowy podział zbioru na typy impulsów. Algorytm klasyfikował część impulsów do błędnych kategorii. Poniżej przedstawiono graficzne zobrazowanie podziału danych na kategorie. Poszczególne zbiory oznaczane są różnymi kolorami. Algorytm podziału działał tutaj na całych wektorach danych wejściowych, co oznacza, że wynikiem klasyfikacji jest umieszczenie zbioru w przestrzeni wielowymiarowej, co powoduje trudność wizualizacji podziału. W celu lepszego zobrazowania rozkładu klas, dane wejściowe metody k-średnich zostały zredukowane do 2 wymiarów z pomocą algorytmu PCA. Proces ten pozwolił na lepsze odwzorowanie prawdziwego rozkładu klas na wykresach dwuwymiarowych.

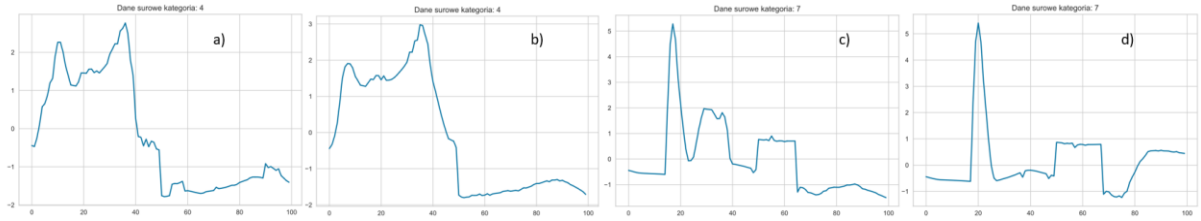


Rysunek 50 Podział danych na 7 kategorii dla a) prądu b) napięcia i c) połączenia prądu i napięcia

Jak widać na powyższym rysunku 50, analiza surowych danych pomiarowych nie pozwala na jednoznaczny podział na kategorie impulsów. Zbiory przenikają się wzajemnie i nie mają wyraźnych granic, co prowadzi w efekcie do częstego klasyfikowania impulsów do złej kategorii. Analizę poprawności klasyfikacji potwierdzano poprzez wybór dużej ilości

losowych przebiegów z każdej kategorii i porównywaniu ich charakteru. Dodatkowo sprawdzano po 1000 kolejnych impulsów z różnych etapów drążenia. Analizując podzielone już na kategorie zestawy danych określono, że zestawy danych łączących przebiegi prądu i napięcia pozwalały na najłatwiejsze interpretowanie wyników klasyfikacji oraz dawały wyniki najbliższe poprawnej klasyfikacji. Z tych powodów dalszej analizie poddawane były właśnie te przebiegi.

Przegląd sklasyfikowanych danych potwierdził, że podział impulsów w oparciu o surowe dane nie daje zadawalających efektów. Algorytm podziału na kategorie skupiał się na podziale różnych typów wyładowań normalnych oraz obwodów otwartych. Wyraźnie zaobserwować można było, że główną cechą braną pod uwagę przy podziale na kategorie była maksymalna amplituda prądu impulsu. Algorytm w większości pomijał inne czynniki takie jak kształt przebiegów. Poniższy rysunek 51 przedstawia przykłady błędnie sklasyfikowanych impulsów.

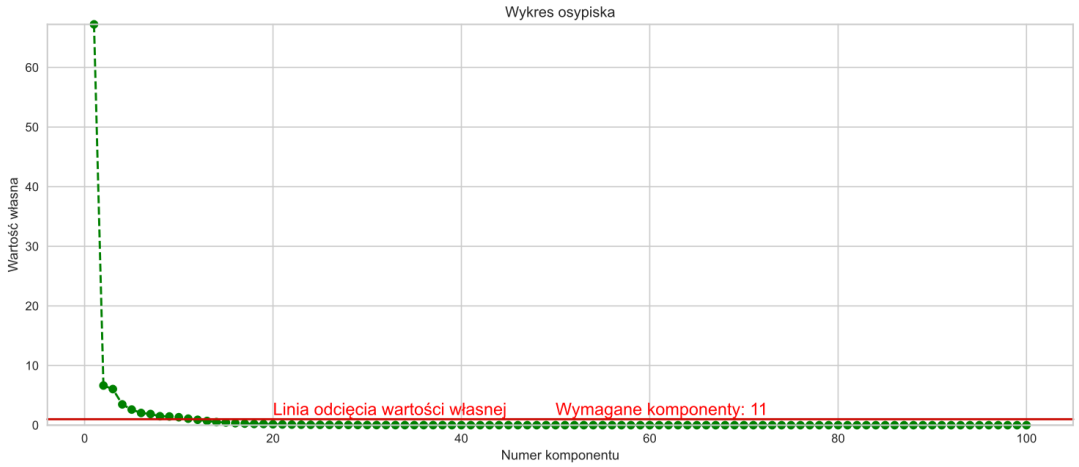


Rysunek 51 Przykładowe przebiegi przydzielone do kategorii na postawie surowych danych

Podstawowym problemem kategoryzowania impulsów, było zaklasyfikowanie wszystkich wyładowań łukowych i zwarć do tej samej kategorii. Rysunek 51 przedstawia przykład takiej klasyfikacji. Wyładowanie łukowe (Rysunek 51 a)) i zwarcie (Rysunek 51 b)) są przypisane do tej samej kategorii o etykiecie nr 4. Rysunek 51 c) i d) obrazuje klasyfikację impulsów o różnych przebiegach, ale dalej sklasyfikowanych do tej samej kategorii.

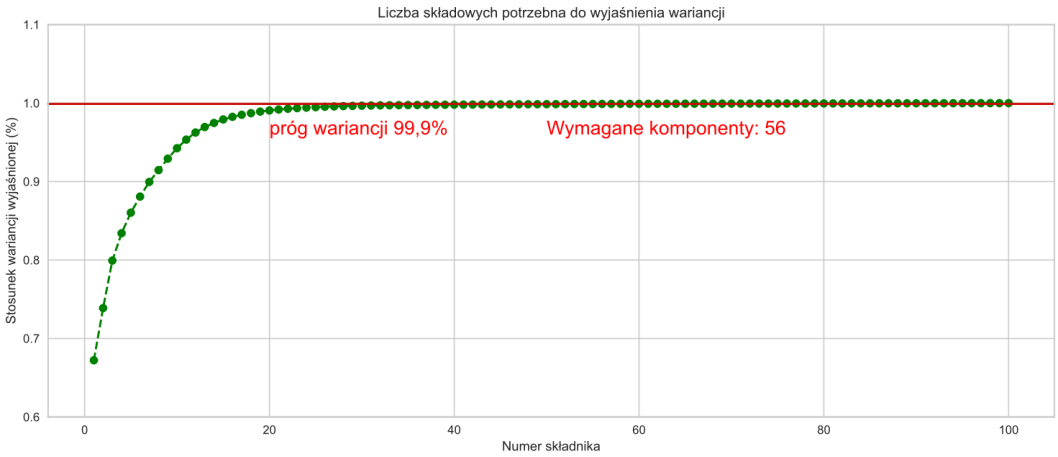
W celu ograniczenia zjawiska błędnej kategoryzacji zastosowano strategie zmniejszenia wymiarowości dla danych wejściowych. Ograniczenie ilości cech powinno pozwolić na wyselekcjonowanie danych zachowujących najbardziej istotne informacje przy jednoczesnym zmniejszaniu złożoności obliczeniowej. Redukcja wymiarowości usuwa nieistotne cechy z danych, ponieważ mogą one obniżyć dokładność algorytmów uczenia maszynowego. Zastosowano podejście oparte o analizę składowych głównych PCA (rozdział 0) oraz oparty o DL Autoenkoder (rozdział 0).

Stosując PCA określono ilość cech do jakiej można ograniczyć dane wejściowe, aby zachować większość istotnych informacji w nich zawartych. W celu oceny poszczególnych wartości własnych zastosowano wykres osypiska (Rysunek 52).



Rysunek 52 Wykres osypiska analizy PCA impulsów

Na poziomej osi wykresu znajdują się numery kolejnych wartości własnych, a na osi pionowej są wartości parametrów własnych. Liczba wartości własnych wynoszących więcej niż jeden, jest równa zaledwie 11. Oznacza to, że możliwa jest znaczna redukcja wymiarowości i już nieduża liczba wartości własnych będzie dobrze reprezentowała dane.



Rysunek 53 Wykres wartości stosunku wariancji wyjaśnionej PCA w funkcji ilości składników dla danych impulsów drążenia

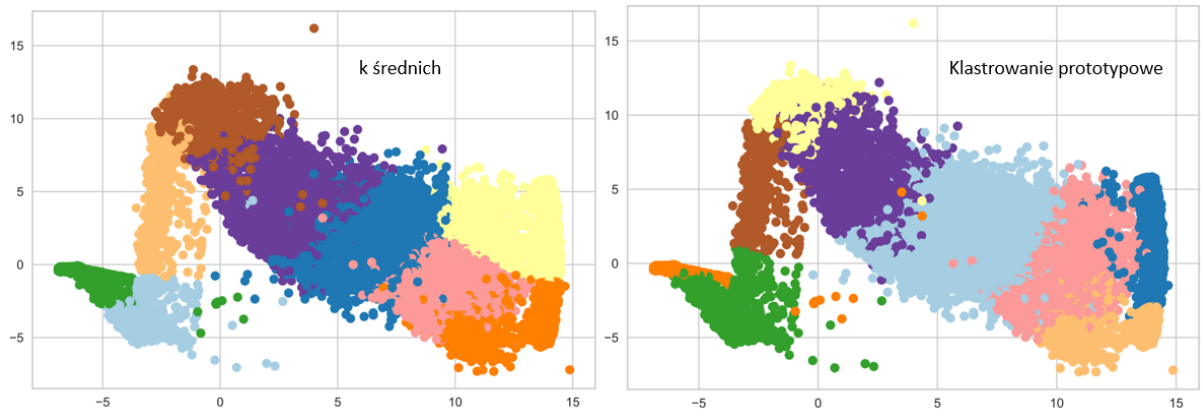
Przeanalizowano także stosunek wariancji wyjaśnionej PCA w zależności od ilości składników głównych. Stosunek wariancji jest uzyskiwany poprzez wektor wartości osobliwych S zgodnie z poniższym równaniem:

$$\frac{\sum_{i=1}^{k} S_i}{\sum_{i=1}^{n} S_i} \tag{26}$$

Gdzie k to liczba składowych (k-wymiarowych), a n to liczba cech. Na rysunku 53 przedstawiono wykres stosunku wariancji wyjaśnionej PCA.

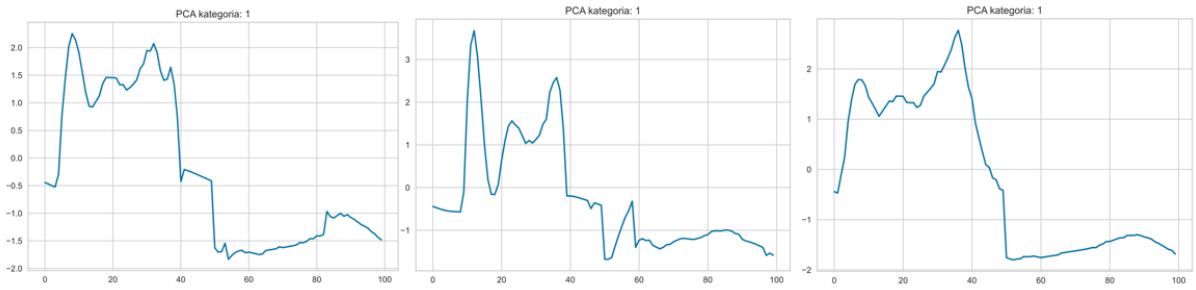
Z analizy wynika, że już 56 składników PCA odpowiada za 99,9% wariancji danych impulsów drążenia. Odpowiednio za 99% wariancji odpowiadało 19 składników, a za 95%, 10 składników. Na podstawie analizy określono, że klasyfikowanie impulsów może zostać wykonane na danych o zmniejszonej wymiarowości praktycznie bez utraty informacji w nich zawartych.

Zestaw danych wejściowych zostały poddane redukcji wymiarowości PCA do 40 cech i następnie podzielone na podgrupy stosując metodę k średnich oraz klastrowanie hierarchiczne (Rysunek 54).



Rysunek 54 Wyniki podziału danych na 9 klas dla analizy k średnich i klastrowania prototypowego zredukowanych metodą PCA danych impulsów

Dane zredukowane metodą PCA (redukcja do 40 składników) pozwalają na znacznie lepszy podział danych na kategorie impulsów. Zastosowanie PCA znacznie poprawiło jakość klasyfikacji, natomiast nie pozwoliło nadal na osiągnięcie w pełni prawidłowej klasyfikacji. Algorytm brał po uwagę kształt przebiegów impulsów, natomiast nadal nie radził sobie w pełni z odróżnianiem wyładowań łukowych i zwarć. Przykład takich impulsów pokazano na poniższym rysunku 55.



Rysunek 55 Błędnie sklasyfikowane wylądowania łukowe i zwarcia po redukcji wymiarowości PCA

Metoda k-średnich i klastrowanie prototypowe dawały podobne wyniki. Przykład klasyfikacji oboma metodami pokazana na rysunku 54. Pomimo tego należy zauważyć, że pomiędzy metodami występowały istotne różnice. Minusem analizy k średnich jest fakt, że każdorazowe uruchomienie procedury może dawać różne i trudne do przewidzenia wyniki, tzn. etykiety klas były różne dla różnych wywołań tego samego algorytmu. Klastrowanie aglomeracyjne nie cechuje się tym ograniczeniem, tzn. analizy dawały powtarzalne wyniki za każdym wywołaniem. Niestety dużym ograniczeniem tej metody jest wysoka złożoność obliczeniowa i konieczność rezerwacji dużej ilości pamięci. Klastrowanie aglomeracyjne wymaga obliczenia i przechowywania macierzy odległości o wymiarach n×n, gdzie n liczba klastrowanych impulsów. W efekcie przetwarzanie dużych ilości danych jest praktycznie niemożliwe. Klasyfikacja częściowych przebiegów drążenia była możliwa, natomiast próba grupowani pełnych danych z pojedynczego procesu drążenia otworu okazała się niemożliwa.

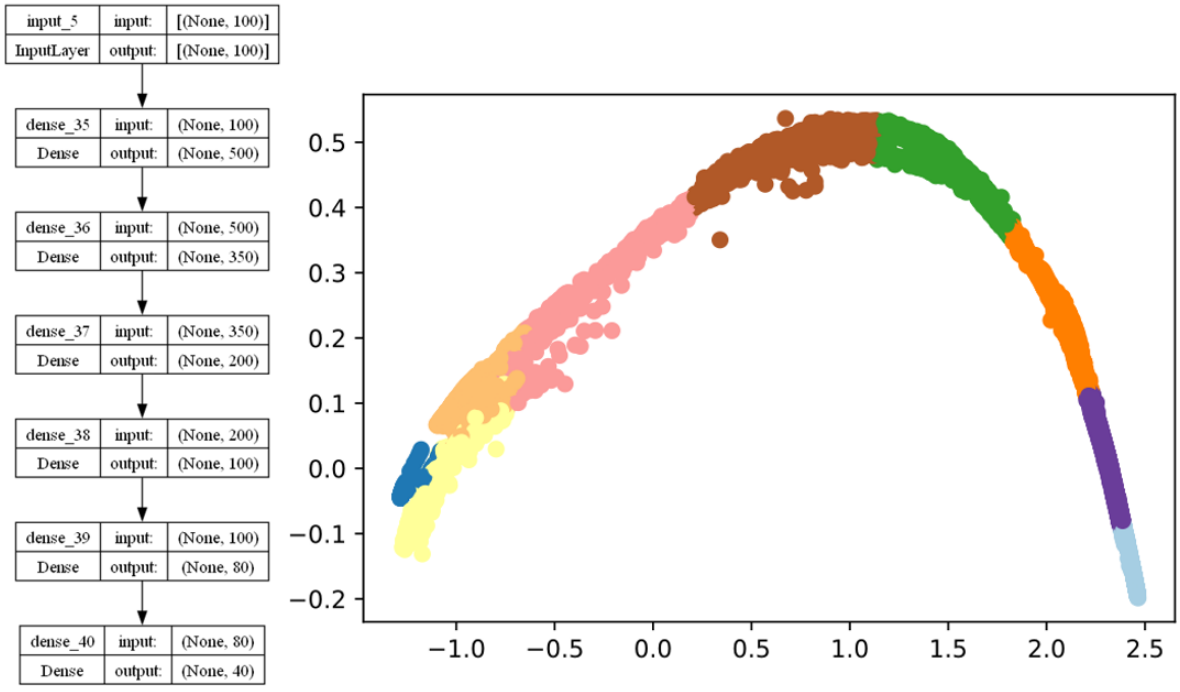
PCA jest fundamentalnie transformacją liniową przez co ma ograniczenia. Ponieważ cechy PCA są projekcjami na ortogonalną bazę, są one całkowicie liniowo nieskorelowane. Ograniczenia te można obejść stosując redukcję wymiarowości danych z pomocą Autoenkodera. Autoenkodery mogą opisywać skomplikowane procesy nieliniowe. Jednakże, ponieważ cechy enkodowane są trenowane tylko do poprawnej rekonstrukcji, mogą występować korelacje. Autoenkodery jak opisywano w rozdziale 0, są sieciami neuronowymi, które mogą redukować dane wejściowe do niskowymiarowej przestrzeni ukrytej poprzez stosowanie wielu przekształceń nieliniowych (warstw). Autoenkoder jest trenowany w procesie propagacji wstecznej poprzez minimalizację różnicy między danymi oryginalnymi, a ich rekonstrukcjami, co umożliwia sieci automatyczne uczenie się najbardziej informatywnej reprezentacji ukrytej.

Po wytrenowaniu pełnego autoenkodera do prawidłowego odwzorowywania danych wejściowych, można zastosować niepełną formę enkodera, jako narzędzie redukcji

wymiarowości. Ponieważ przestrzeń ukryta ma niższe wymiary niż wejście, zmienne ukryte enkodera będą kodować najważniejsze cechy wejścia, aby sieć dekodera była zdolna do jego rekonstrukcji.

Co prawda PCA jest szybsze i mniej kosztowne obliczeniowo niż autoenkodery, natomiast w przypadku dużych zbiorów danych, które nie mogą być przechowywane w pamięci głównej, PCA nie może być zastosowane. W takim przypadku autoenkodery mogą być stosowane, ponieważ mogą pracować na partii danych o mniejszych rozmiarach i dlatego ograniczenia pamięci nie wpływają na redukcję wymiarów za pomocą autoenkoderów. W wypadku dużych ilości danych wysokoczęstotliwościowych procesu EDM jest to duża zaleta.

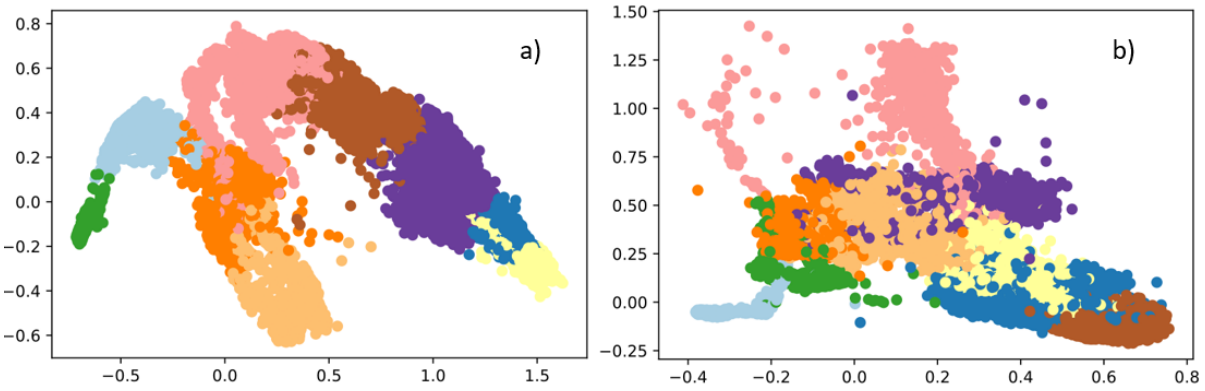
W niniejszej pracy zaproponowano zbudowanie głębokiego przepełnionego i niepełnego enkodera redukującego dane. Przetestowane zostały różne kombinacje głębokości, ilości połączeń ukrytych, funkcji aktywacyjnych oraz algorytmów optymizacji i funkcji straty. Dopiero zastosowanie głębokiego autoenkodera o dużej ilości połączeń ukrytych pozwoliło na istotną poprawę automatycznej klasyfikacji impulsów.



Rysunek 56 Budowa enkodera redukującego dane oraz graficzne przedstawianie klastrowania danych wejściowych zredukowanych z jego pomocą (k=9)

Na powyższym rysunku 56 zobrazowano strukturę końcowego enkodera redukującego wymiarowość (po lewej) oraz zobrazowanie przykładowego podziału danych na 9 kategorii impulsów.

W celu szkolenia zbudowano pełny, wielowarstwowy autoenkoder złożony zarówno z części enkodera jak i symetrycznej części dekodera. Optymalne wyniki osiągano stosując przepełnioną warstwę wejściową, dalej 4 warstwy ukryte o zmniejszającej się ilości neuronów z funkcją aktywacji ReLU (ang. Rectified Linear Unit), warstwie „wąskiego gardła" z 40 połączeniami i sigmoidalną funkcją aktywacji i symetrycznym do części enkodera dekoderem. Stosowanie sieci o mniejszej ilości warstw ukrytych i podejścia niedopełnionego (tylko zmniejszanie wymiarowości poniżej wymiarowości danych wejściowych), owocowało niejednoznacznymi wynikami klasyfikacji. Przykład takiej klasyfikacji pokazano na poniższym rysunku 57.



Rysunek 57 Wyniki klasyfikacji danych zredukowanych enkoderami niedopełnionymi a) 3 warstwy ukryte b) 2 warstwy ukryte

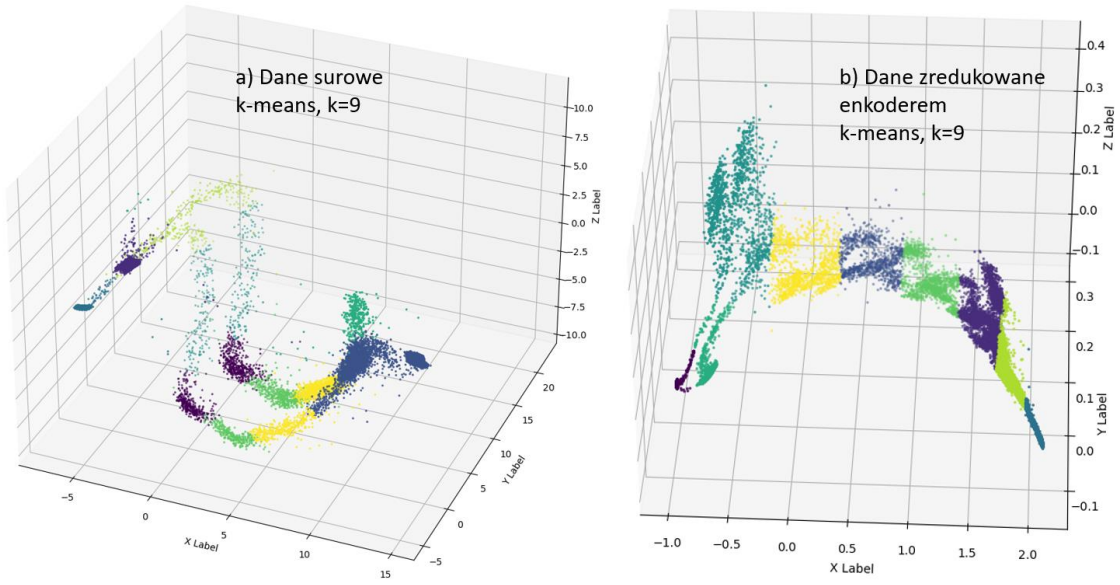
Tak zbudowaną sieć poddano uczeniu nadzorowanemu. Zadaniem pełnego autoenkodera było jak najlepsze odwzorowanie danych wejściowych do tej samej wymiarowości co dane wejściowe. Optymalne wyniki klastrowania osiągano dla funkcji straty binarnej entropii krzyżowej oraz optymalizatora AdaDelta [315]. AdaDelta to technika optymalizacji stochastycznej, która pozwala na ustalanie współczynnika uczenia się na poziomie wymiaru dla SGD. Jest to rozwinięcie metody Adagrad, której celem jest zmniejszenie agresywnego, monotonicznie malejącego współczynnika uczenia się. Zamiast gromadzić wszystkie poprzednie kwadraty gradientów, AdaDelta ogranicza okno akumulowanych poprzednich gradientów do ustalonego rozmiaru. Stosowanie innych optymalizatorów jak algorytm Adam [301] i standardowe SGD, oraz bardziej standardowych w tego typu zagadnieniach funkcji strat jak np. błędy średniokwadratowe pozwalało na osiąganie bardzo małych wartości funkcji straty

co przekładało się na bardzo dobre odtwarzanie danych wejściowych przez autoenkoder. Co ciekawe, zastosowanie tak wytrenowanego niepełnego enkodera do procesu podziału na grupy impulsów nie przynosiło oczekiwanej poprawy jakości klastrowania. Dopiero stosowanie kombinacji przepełnionego enkodera, optymalizatora AdaDelta i binarnej krosentropii dla danych standaryzowanych przynosiło wyraźną poprawę jakości klastrowania. Wynikało to najprawdopodobniej z faktu, że stosowanie takiego podejścia wprowadzało dodatkową transformację danych wejściowych zamiast samego jak najlepszego odwzorowywania wszystkich istotnych cech sygnału.

W procesie uczenia ze względu na stosowanie kombinacji standaryzacji i binarnej entropii krzyżowej zaobserwowano wyraźną tendencję autoenkodera do przeuczania. Funkcja straty nie była zbierzna do zera i mogła przyjmować ujemne wartości. Przy zbyt dużej ilości epok uczenia następowała degradacja wyników klasyfikacji przeprowadzanej na danych zredukowanych enkoderem. Aby temu zapobiec wprowadzono warunek automatycznego zatrzymania procesu uczenia, na epoce dającej najmniejszą wartość funkcji straty poprzez odpowiedni dobór ilości epok oraz ilości wielkość partii uczącej w epoce.

Z nauczonego pełnego enkodera, wyodrębniono samą cześć kodującą, czyli warstwę wejściową, warstwy ukryte oraz warstwę „wąskiego gardła” sieci. W efekcie dane wejściowe przetworzone enkoderem były modyfikowane i redukowane do wymiarowości o 40 cechach. Jak pokazała analiza stosunku wariancji wyjściowej w zależności od ilości cech, taka ilość cech zachowywała około 99% informacji z danych wejściowych. Możliwe było dalsze zmniejszanie ilości cech zredukowanych, natomiast badania wyników klasyfikacji prowadzonej na podstawie taki danych pokazywały, że powodowało to zmniejszenie dokładności.

Dane przetworzone tak stworzonym enkoderem posłużyły dalej do procesu grupowania na zestawy kategorii impulsów. Aby lepiej zobrazować różnice w grupowaniu wprowadzone przez zaproponowany enkoder, dokonano klastrowania zestawów danych surowych oraz danych zredukowanych, na 9 kategorii oraz zmniejszono wymiarowości obu wariantów danych do 3 cech z użyciem PCA. Taka wymiarowość danych pozwoliła na zobrazowanie rozkładu kategorii impulsów dla całej populacji na wykresach 3d.

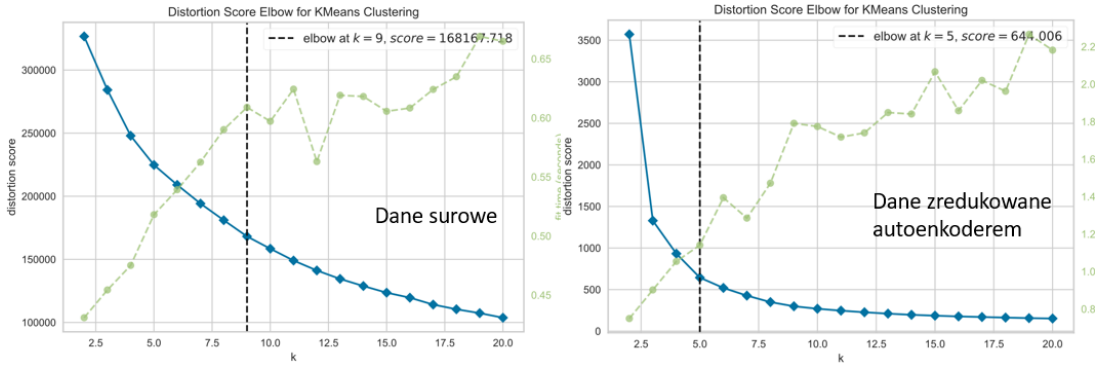


Rysunek 58 Rozkład kategorii impulsów dla klastrowania danych surowych i przetworzonych enkoderem. Wizualizacja danych wejściowych zredukowanych do 3 cech PCA

Jak pokazuje powyższa wizualizacja na wykresach 3D, przetwarzanie danych zaproponowanym enkoderem powoduje ich wyraźne przekształcenie. Dane rzutowane są na pewną hiperpłaszczyznę. Operacja ta pozwala na bardziej jednoznaczne dzielenie zestawu danych na grupy. W wypadku danych surowych (Rysunek 58 a)) widać złożony charakter rozkładu danych, nawet na projekcji do 3 wymiarów. Zaobserwować można także dużą ilość punktów znacząco oddalonych od centroidów poszczególnych grup. Punkty odstające często występują w sąsiedztwie punktów klasyfikowanych do innych grup, co może być przyczyną błędnych klasyfikacji impulsów. W wypadku danych przetworzonych enkoderem (Rysunek 58 b)) poszczególne kategorie są znacznie bardziej zgrupowane. Obserwowane pojedyncze punkty odstające nie występują w bliskiej odległości od punktów innych grup i są grupowane jednoznacznie do danej kategorii. Testy poprawności grupowania impulsów na takich danych potwierdziły lepszą jakość klasyfikacji.

Należy zadać sobie pytanie na ile klas dzielić impulsy. Analiza literaturowa i przegląd przebiegów opisany w rozdziale 4.5 sugeruje, że minimalną liczba etykiet powinna pozwalać na odróżnianie impulsów normalnych, wyładowań łukowych, obwodów otwartych i zwarć. Jak pokazano w niniejszej pracy, impulsy klasyfikowane jako normalne wykazują dużą zmienność. Mogą posiadać różne amplitudy, występować z różnym opóźnieniem i posiadać różne przebiegi. Z tych względów próba podziału impulsów normalnych na dodatkowe podkategorie będzie zasadna i może dostarczać dodatkowych informacji w analizie procesu.

Podstawową ideą metod klastrowania, jest zdefiniowanie klastrów w taki sposób, aby całkowita wewnątrzgrupowa zmienność [lub całkowita suma kwadratów wewnątrzgrupowych (WSS)] była zminimalizowana. W celu określenia ilości klastrów minimalizującą tą odległość zastosowano metodę krzywej łokcia [317]. Metoda ta przeprowadza klastrowanie k-średnich na zbiorze danych dla różnych wartości k (liczba klastrów). Dla każdej z wartości k obliczono średnie odległości do centroidy dla wszystkich punktów danych. Następnie sporządzono wykres tych średnich odległości w funkcji ilości klastrów i określono punkt, w którym średnia odległość od centroidów gwałtownie spada („Łokieć").



Rysunek 59 Wykresy łokcia dla danych klastrowania danych surowych i danych zredukowanych autoenkoderem

Analizując wykresy krzywych łokcia dla danych surowych otrzymaliśmy wartość 9 kategorii jako optymalną, w danych zredukowanych z pomocą autoenkodera otrzymaliśmy wartość k=5 jako optymalną.

Dalej zastosowano analizę sylwetki (ang. Silhouette Analysis) [318]. Współczynnik sylwetkowy lub wynik sylwetki dla k-means to miara tego, jak podobny jest punkt danych wewnątrz klastra (spójność) w porównaniu z innymi klastrami (separacja). Współczynnik sylwetkowy dla konkretnego punktu danych określony jest wzorem:

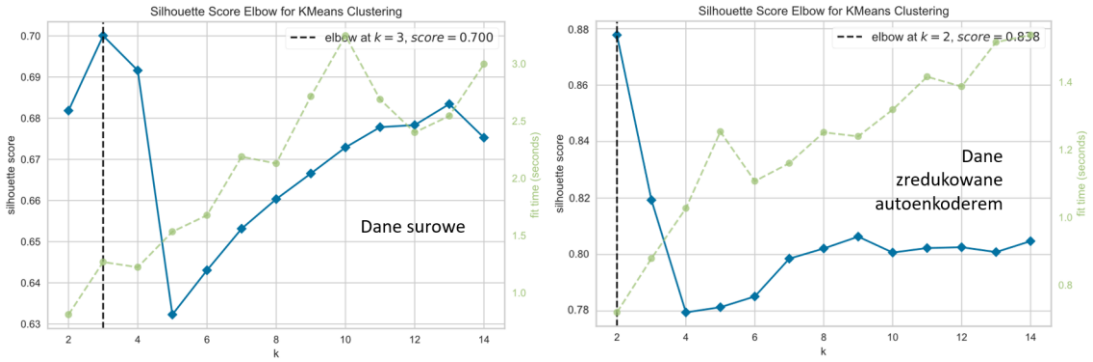
$$S_i = \frac{b_i - a_i}{\max\{a_i, b_i\}} \tag{27}$$

Gdzie S_i to współczynnik sylwetkowy punktu danych i, a_i to średnia odległość między punktem i, a wszystkimi innymi punktami danych w klastrze, do którego należy i, b_i to średnia odległość od i do wszystkich klastrów, do których punkt i nie należy.

Następnie obliczana jest średnia wartość współczynnika sylwetki wg wzoru:

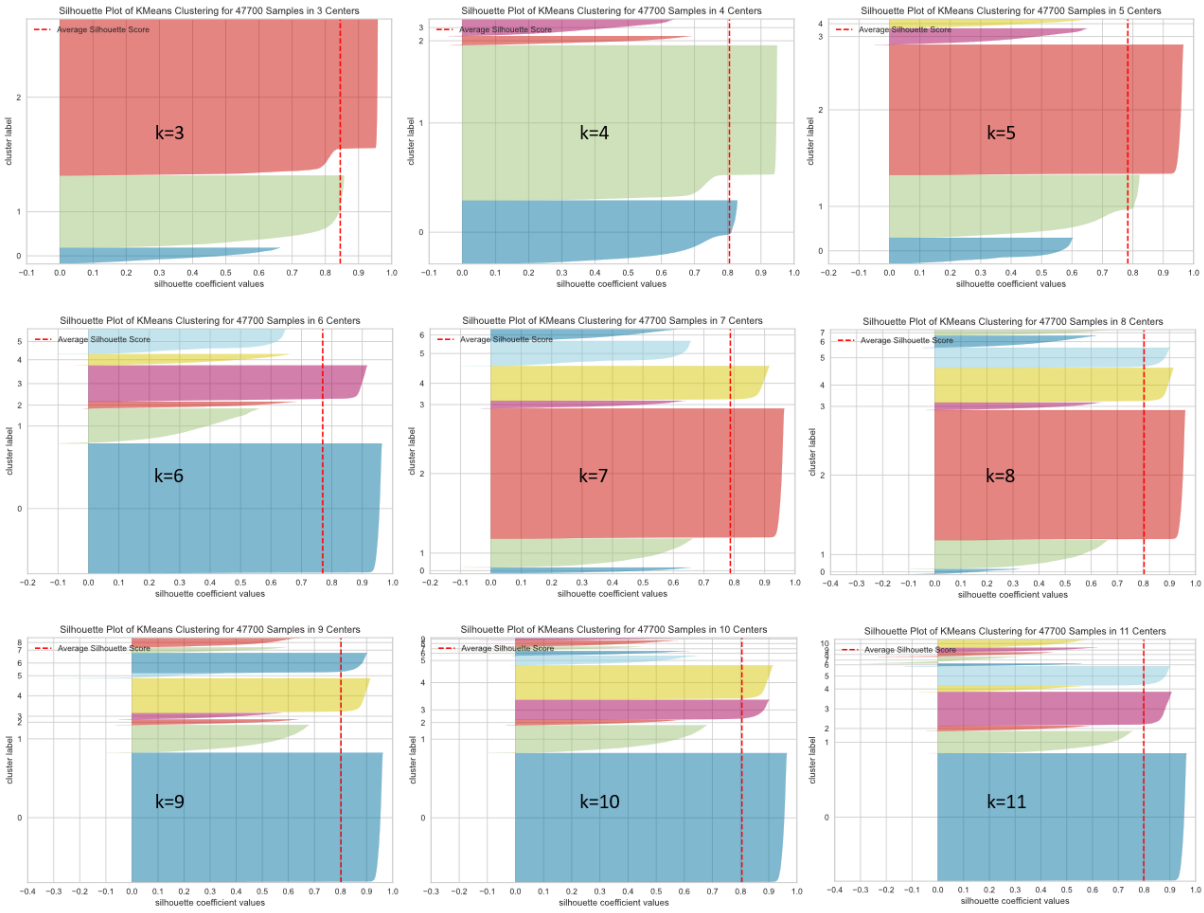
$$\bar{S}=\text{mean}\{S_i\} \tag{28}$$

Dalej wykreślono wykres średniej wartości współczynnika sylwetki w funkcji ilości klastrów.



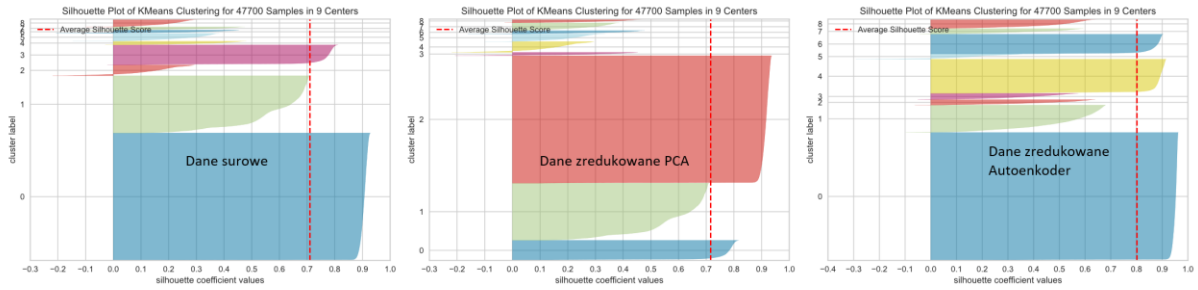
Rysunek 60 Wykres współczynnika sylwetki w funkcji ilości klastrów dla danych surowych i danych zredukowanych enkoderem

Analizując wykres współczynnika sylwetki można dojść do wniosku, że optymalna będzie mała ilość kategorii impulsów (k=2 lub k=3). Z punktu widzenia analizy procesu jest to założenie błędne. Istotne jest dobranie ilości podzbiorów, która pozwoli na odróżnianie podstawowych typów impulsów EDM. Z tego względu wartość k=9 wydaje się optymalna. Wartość współczynnika sylwetki osiągną dla niej lokalne maksimum. Taka ilość zbiorów powinna dawać też dostateczną możliwość rozróżniania zasadniczych typów impulsów opisywanych w literaturze [49-52] oraz zaobserwowanych w danych czasowych (rozdział 4.5). Dopiero głębsza analiza wykresów sylwetki dla poszczególnych wartości współczynnika k ujawnia z czego wynika występowanie maksymalnej wartości współczynnika sylwetki dla małej ilości klastrów.



Rysunek 61 Wykres dopasowania punktów z użyciem miary sylwetki dla różnych ilości klastrów (od k=3 do k=11)

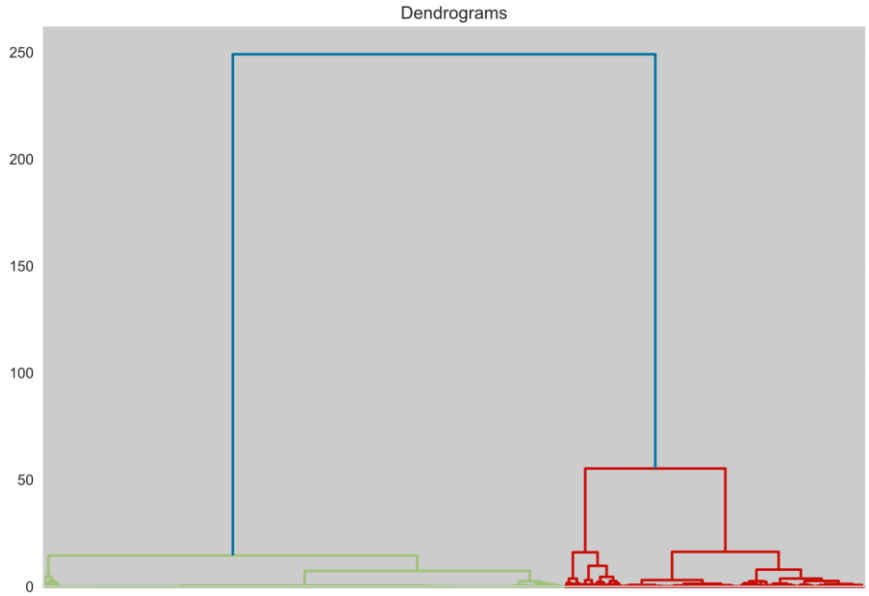
Na powyższym rysunku 61 przedstawiono wykresy dopasowania z użyciem miary sylwetki dla różnej ilości klastrów (od k=3 do k=11). Jak widać, średnia wartość współczynnika sylwetki osiąga największą wartość dla małej ilości klastrów. Zwiększanie ilości klastrów prowadzi do zmniejszenia wartości średniego współczynnika. Wynika to z charakteru klasyfikowanych impulsów. Największy liczebnie klaster danych odpowiada impulsom otwartych obwodów. Zwiększenie ilości kategorii dzieli okresy, w których występowały różne formy impulsów na podkategorie, bez zmniejszania kategorii najbardziej liczebnej. Co za tym idzie ilość klastrów powinna zostać dobrana tak, aby zapewnić dostateczne rozróżnienie interesujących typów impulsów. Analizując charakter sklastrowanych impulsów ustalono, że k=9 daje wystarczające zróżnicowanie impulsów. Zwiększanie ilości kategorii prowadziło do tworzenia nowych klastrów o małej liczebności próbek i o charakterze impulsów zbliżonym do innych kategorii, co pogarszało dalej jakość podziału. Dane podzielone na 9 kategorii pozwoliły na dalszą analizę przebiegu procesu.



Rysunek 62 Porównanie dopasowania punktów do klastrów na podstawie miary sylwetki dla danych surowych i zredukowanych z użycie PCA i autoenkodera

Powyższy rysunek 62 przedstawia porównanie wykresów dopasowania punktów dla danych surowych i danych zredukowanych z użyciem PCA i enkodera. Podział na kategorie dla danych surowych i PCA są zbliżone, natomiast można zaobserwować różnice liczebności klastrów autoenkodera. Wszystkie metody w podobny sposób klasyfikowały najbardziej liczne impulsy otwarte, natomiast różnice występowały w podziale danych zawierających różne typy impulsów EDM. Należy zwrócić również uwagę na lewą stronę wykresów, gdzie współczynniki sylwetki przyjmują wartości ujemne. Duża ilość wartości ujemnych wskazuje na brak wyraźnej różnicy między obiektami jednej klasy, a (przynajmniej jedną) inną klasą. Fakt ten prowadził do błędnego klasyfikowania impulsów. W wypadku danych zredukowanych autoenkoderem zaobserwowano mało ujemnych współczynników sylwetki pozwalając na prawidłowe klasyfikowanie impulsów, co zostało potwierdzone przeglądem sklasyfikowanych impulsów z różnych kategorii.

Podział na kategorie impulsów potwierdzony został także z użyciem wykresu dendrogramu (Rysunek 63) tworzonego z użyciem klastrowania hierarchicznego [312]. Na pionowej osi dendrogramu obrazowana jest odległość pomiędzy klastrami. Zaobserwować można, że istnieje niewielka liczba klastrów o dużej odległości między sobą oraz potencjalnie duża ilości klastrów o dużym podobieństwie. Odpowiada to podziałowi danych na impulsy obwodów otwartych, impulsy zwarć i wyładowań łukowych i wyładowań normalnych. Każda z tych trzech grup wykazuje zasadnicze różnice między sobą co widać po wysokości dendrogramu. Ponieważ impulsy normalne wykazują dużą zmienność tą grupę można podzielić na dużą ilość klastrów o niewielkich odległościach między sobą, co jest obrazowane na dendrogramie kolorem czerwonym.



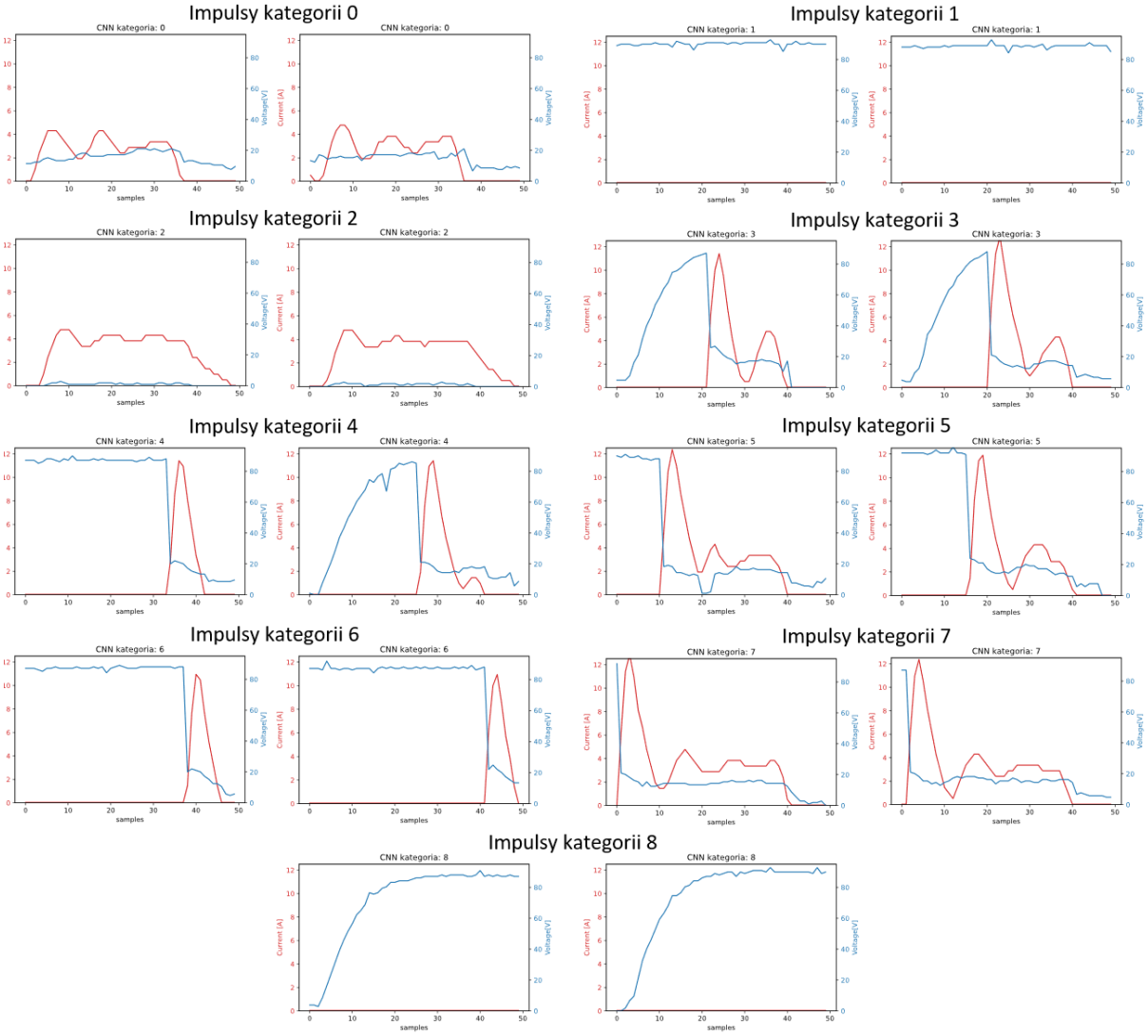
Rysunek 63 Dendrogram procesu klastrowania impulsów

Oprócz opisywanego powyżej automatycznego klastrowania na określoną ilość kategorii wdrożono także dodatkowe rozwiązanie programowe pozwalające na sterownie procesem podziału.

Podejście opierało się na wielostopniowym procesie podziału zbioru danych na określoną ilość grup. Po każdym z podziałów użytkownik miał możliwość przeglądu losowo wybranych lub określonej ilości kolejno występujących impulsów danej kategorii w wybranym obszarze przebiegu. Na każdym kroku użytkownik miał możliwość nadania danej grupie impulsów określonej etykiety. Jeśli dana grupa wykazywała zmienność, możliwe było podzielenie grupy na kolejną ilość podgrup i ponowne etykietowanie poprzez przydzielenie ich do grup istniejących lub stworzenie nowych. Podejście takie pozwalało na osiąganie podobnie wysokiej dokładności klasyfikacji impulsów jak w wypadku podziału automatycznego z enkoderem. Zastosowanie obu podejść dawało możliwość przydzielania grupom impulsów tych samych etykiet w analizie różnych zestawów danych.

Ostatecznie procedury opisywane powyżej pozwoliły na prawidłowy podział zestawów danych odpowiadających poszczególnym okresom EDM drążenia na 9 grup. Zastosowane metody pozwoliły na osiągnięcie subiektywnej dokładności (określanej opisywanymi wcześniej metodami losowania i przeglądu konsekutywnych impulsów) klasyfikacji na poziomie przekraczającym 99% poprawności.

Jak pokazano na rysunku 64, algorytm wyodrębnił impulsy przedstawiające kilka grup charakteryzujących się konkretnymi przebiegach napięcia i prądu. Automatyczny podział na grupy pozwolił na poprawne zidentyfikowanie podstawowych kategorii impulsów rozpoznanych w wyniku przeglądu szeregów czasowych procesu (rozdział 4.5).



Rysunek 64 Przykładowe przebiegi impulsów spośród 9 wyodrębnionych kategorii (parametry #1)

Wyszczególnione kategorie impulsów to:

- Kategoria 0 – wyładowania łukowe,
- Kategoria 1 – obwody otwarte. Brak generacji iskry w okresie *T*, napięcie utrzymywane na maksymalnej wartości, co oznacza brak generacji impulsu w poprzednim cyklu,
- Kategoria 2 – wyładowania zwarciowe,

- Kategoria 3 – Impuls podwójny występujący z opóźnieniem T_d,
- Kategoria 4 – Impuls pojedynczy występujący z opóźnieniem T_d,
- Kategoria 5 - Impuls występujący z opóźnieniem T_d oraz podtrzymaniem przepływy prądu do końca okresu T_{on},
- Kategoria 6 – Impuls pojedynczy występujący na końcu okresu EDM,
- Kategoria 7 – impulsy prądowe zachodzące od początku okresu T z utrzymującym się do końca czasu T_{on} przepływem prądu o mniejszej amplitudzie,
- Kategoria 8 - obwód otwarty z narastaniem poziomu napięcia w czasie T_{on}, co oznacza wystąpienie impulsu w poprzednim cyklu.

Opracowana metoda pozwoliła na podział impulsów prądowych na istotne z punktu widzenia ich charakterystyki grupy. Podział pozwolił na identyfikację zasadniczych grup impulsów opisywanych w literaturze [49-52] i publikacjach opisujących metody identyfikacji przebicia FHD [63-69]. Ciekawa natomiast jest duża zmienność impulsów. Źródła literaturowe w większości opisują kilka idealnych przebiegów impulsów. W prowadzonych badaniach, szczególnie impulsy normalne, miały zróżnicowane przebiegi. Wyjątkowo interesująca jest także obserwowana kategoria impulsów nr 6. W źródłach, opisywany czas cyklu T_{off} nazywany jest czasem bezimpulsowym. W tym okresie, ideowo, impulsy nie powinny zachodzić. Natomiast, jak wykazała analiza, budowa maszyny MKG 1000 pozwala na wystąpienie takich impulsów. W wypadku niezaistnienia impulsu w danym cyklu EDM, w kolejnym cyklu napięcie utrzymywane jest na maksymalnym ustawionym poziomie. Jak widać, przy jednoczesnym zaistnieniu odpowiednich warunków procesu, sytuacja taka pozwala na wygenerowanie wyładowania w okresie T_{off}.

Podobny proces klastrowania przeprowadzono dla grupy drążeń wykonywanych z parametrami agresywnymi (parametry #2). Dla parametrów optymalnych i agresywnych charakter impulsów był podobny z zauważalną zmienną ilością wystąpień różnych typów.

5.1.3. Narzędzie klasyfikowania impulsów

Niestety, jak opisywano w poprzednim rozdziale, wielokrotne wywoływanie algorytmu k-średnich może dawać różne wyniki grupowania, oraz posiada ograniczenia ze względu na zajmowanie dużej ilości pamięci, co uniemożliwia bezpośrednie zastosowanie tej metody dla danych zarejestrowanych w różnych drążeniach (różnych plików z danymi). Aby zapobiec temu ograniczeniu, opracowano klasyfikator oparty o głębokie sieci neuronowe.

Zestawy danych z różnych drążeń zostały wykorzystane do nadzorowanego nauczenia klasyfikatora impulsów. Tak przygotowane narzędzie klasyfikacji mogło być użyte wielokrotnie do przetwarzania danych ze wszystkich drążeń, dając powtarzalne i interpretowalne wyniki. Tworząc narzędzie klasyfikacji przetestowano podejścia oparte o klasyczne uczenie maszynowe: regresję logistyczną, SVM z jądrem funkcji liniowej i radialnej RBF (rozdział 0), drzew decyzyjnych (rozdział 0), lasów losowych oraz metodę kNN (rozdział 0). Dodatkowo przetestowano kilka podejść opartych o uczenie głębokie: Sztuczne sieci neuronowe MLP (rozdział 0), konwolucyjne sieci neuronowe CNN 1D (rozdział 0), rekurencyjne sieci neuronowe RNN i LSTM (rozdział 0), oraz sieci głębokich przekonań DBN (rozdział 0).

W celu wyboru optymalnego modelu klasyfikacyjnego, w procesie uczenia każdej z metod stosowano metody przeszukiwania siatki (ang. grid search) oraz przeszukiwanie losowe (ang. random search) do optymalizacji hiperparametrów modelu oraz metody wydzielania (ang. holdout cross-validation) i k-krotną walidację krzyżową (k-fold cross-validation) w celu oceny jakości klasyfikatora.

Metody przeszukiwania pozwoliły na zautomatyzowanie doboru optymalnych parametrów stosowanych metod klasyfikacji. Przeszukiwanie siatki polegało na systematycznym sprawdzaniu skuteczności modeli dla wszystkich określonych zestawów hiperparametrów. Dodatkowo stosowano przeszukiwanie losowe, które sprawdzało losowe kombinacje parametrów modeli z określonych przedziałów. Takie samo podejście zastosowane zostało także w procesie optymalizacji stosowanych głębokich sieci neuronowych zmieniając parametry takie jak np. głębokości sieci, ilości połączeń, funkcje aktywacyjne, czy ilości filtrów itp. w zależności od metody. W celu oceny wydajności każdego z wariantów klasyfikatorów zastosowano dwie metody: wydzielanie i k-krotną walidację krzyżową. Przygotowane dane,

zgodnie z ideą wydzielania, zostały podzielone na dwie grupy: dane uczące stanowiące 70% losowo wybranych danych z populacji oraz dane testowe stanowiące pozostałe 30%. Do danych testowych oprócz losowo wybranych impulsów zaliczono także zestawy danych z całych przebiegów drążeń. Podejście takie miało na celu sprawdzenie nie tylko jakość klasyfikacji w obrębie całej populacji testowej, ale także dokładność klasyfikacji w obrębie wykonywania pojedynczego otworu. Dane uczące stosowano do procesu nadzorowanego uczenia klasyfikatorów, a dane testowe posłużyły do oceny pełnego modelu. Podczas uczenia stosowano także metodę k-krotnego sprawdzianu krzyżowego [319]. Dostępny zestaw uczący był dzielony na k dysjunktywnych podzbiorów o zbliżonej wielkości. Ponieważ w danych uczących występują dysproporcje między klasami zastosowano podejście warstwowe (stratyfikowane) [320]. Podział na podzbiory prowadzony był przez próbkowanie losowe z zachowaniem proporcji, co oznacza, że próbkowanie jest wykonywane w taki sposób, aby proporcje klas w poszczególnych podzbiorach odzwierciedlały proporcje w zestawie uczącym. Dzięki temu stosunek klas w zbiorze uczącym jest najlepszym oszacowaniem stosunku klas w populacji, co pozwala na uniknięcie błędnego uczenia (biassed learning). Jako ilość podzbiorów zastosowano k=10 jak rekomendują w swojej pracy Kohavi et al. [320]. Model był trenowany przy użyciu 9 (k−1) podzbiorów, które razem tworzą zbiór treningowy. Następnie model był stosowany do pozostałego podzbioru walidacyjnego i mierzona była jego wydajność. Ten proces powtarzany był, aż każdy z dziesięciu podzbiorów został użyty jako zbiór walidacyjny. Średnia z k pomiarów wydajności na k zbiorach walidacyjnych stanowiła wynikową wydajność walidacji krzyżowej. Ponieważ k-krotna kroswalidacja jest procesem próbkowania bez zwracania, oszacowanie skuteczności modelu jest obarczone mniejszą wariancją niż w samej metodzie wydzielania. Podejście to pozwala na ocenę każdego ze sprawdzanych wariantów hiperparametrów, w celu wyboru parametrów optymalnych zapewniających wystarczającą wydajność uogólniania. Po doborze optymalnego modelu, uczony on był na całym zestawie testowym oraz określana była ostateczna skuteczność modelu z pomocą danych testowych przygotowanych w metodzie wydzielania.

Jednym z zagadnień klasyfikacji był wybór danych wejściowych. Jak w wypadku opisywanego w rozdziale 5.1.2 procesu klastrowania, sprawdzono dokładność klasyfikacji różnymi metodami korzystając z przebiegów samego napięcia, samego prądu lub kombinowanych przebiegów złożonych z danych napięcia i prądu. Dodatkowo przeanalizowano wpływ normalizacji i standaryzacji danych wejściowych na dokładność

klasyfikacji. Do porównania wykonano proces klasyfikacji standardowymi metodami ML oraz przykładową implementacją głębokiej konwolucyjnej sieci CNN. Poniższa tabela zestawia wyniki osiągnięte na danych walidacyjnych po uczeniu każdej z metod dla różnych zestawów danych wejściowych.

Tabela 8 Wyniki klasyfikacji impulsów dla różnych danych wejściowych: napięcia, prądu i napięcia+prądu

Metoda klasyfikacji	Dokładność na danych walidacyjnych dla napięcia [%]			Dokładność na danych walidacyjnych dla prądu [%]			Dokładność na danych walidacyjnych dla napięcia+prądu [%]		
	Dane surowe	Dane normalizowane	Dane standaryzowane	Dane surowe	Dane normalizowane	Dane standaryzowane	Dane surowe	Dane normalizowane	Dane standaryzowane
Regresja logistyczna	73,691	91,592	88,248	79,886	89,796	90,177	93,684	97,015	95,341
SVM	-	93,010	93,130	-	91,137	91,318	98,771	98,436	98,810
kNN	96,019	96,000	96,040	94,463	94,476	94,316	97,124	97,583	97,408
Drzewo decyzyjne	95,116	95,114	95,120	93,221	93,223	93,223	96,615	96,600	96,608
Las losowy	96,581	96,571	96,577	94,552	94,577	94,573	97,829	97,825	97,834
CNN	96,460	95,877	95,734	93,854	94,745	94,537	97,124	98,217	98,063

Jak pokazuje powyższa analiza, najlepszą dokładność klasyfikacji niezależnie od metody osiągano stosując kombinowane przebiegi napięcia i prądu. Porównując wyniki klasyfikacji ze względu na dane surowe, normalizowane lub standaryzowane można zauważyć wyższość klasyfikacji na danych przetworzonych. Najlepsze wyniki osiągnięto z użyciem standaryzowanych danych napięcia+prądu z wykorzystaniem metody SVM oraz danych normalizowanych dla sieci neuronowych. Dlatego w kolejnych badaniach ograniczono analizę metod tylko dla danych kombinowanych przebiegów napięcia i prądu.

Różne metody klasyfikacji dość dobrze radziły sobie już na podstawie znormalizowanych lub standaryzowanych danych. Na tym etapie nie była konieczna dodatkowa ekstrakcja cech lub stosowane wcześniej metody redukcji wymiarowości danych wejściowych.

W celu doboru optymalnego narzędzia do klasyfikowania impulsów dokonano serii prób uczenia stosując różne metody, wartości parametrów algorytmów i architektury sieci neuronowych stosując metody optymalizacji opisywane powyżej. Zestawienie wyników przykładowych konfiguracji testowych pokazano w Tabela 9.

Tabela 9 Wyniki klasyfikacji impulsów różnymi metodami

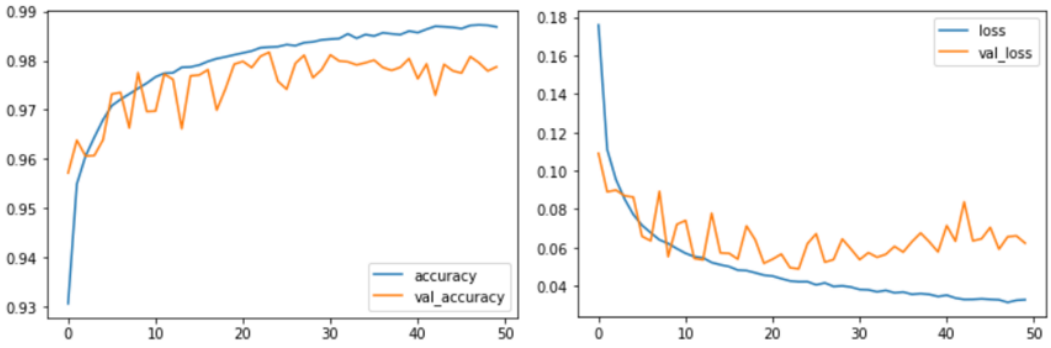
Metoda klasyfikacji	Dokładność na danych walidacyjnych [%]
Regresja logistyczna (odwrotność siły regularyzacji=1, kara=l2)	95,352
SVM (kernel=linear, C=2)	98,817
SVM (kernel=rbf, C=2)	98,86
kNN (9 sąsiadów, metryka minkowskiego, parametr=l2)	97,408
Drzewo decyzyjne (kryterium=gini, głębokość=10)	95,937
Drzewo decyzyjne (kryterium=gini, głębokość=20)	96,608
Drzewo decyzyjne (kryterium=gini, głębokość=50)	96,604
Las losowy (kryterium=gini, ilość drzew=10)	97,638
Las losowy (kryterium=gini, ilość drzew=50)	97,987
Las losowy (kryterium=gini, ilość drzew=200)	98,032
Las losowy (kryterium=gini, ilość drzew=512)	98,029
CNN 1D - input (100,1 ReLU), Conv1D(1x100x16x3 ReLU), maxPool, flatten, dropout(0,2), dense(1600), output(9,softmax)	97,309
CNN 1D - input (100,1 ReLU), Conv1D(1x100x16x3 ReLU), maxPool, Conv1D(1x50x32x3 ReLU), flatten, dropout(0,2), dense(480), output(9,softmax)	97,714
CNN 1D - input (100,1 ReLU), Conv1D(1x100x32x3 ReLU), maxPool, Conv1D(1x50x64x3 ReLU), maxPool, Conv1D(1x25x128x3 ReLU), flatten, dense(480), dropout(0,2), output(9,softmax)	98,334
CNN 1D - input (100,1 ReLU), Conv1D(1x100x16x2 ReLU), maxPool, Conv1D(1x50x32x2 ReLU), maxPool, Conv1D(1x25x48x2 ReLU), maxPool, Conv1D(1x12x64x2 ReLU), maxPool, Conv1D(1x6x80x2 ReLU), flatten, dense(480), dropout(0,2), output(9,softmax)	97,69
CNN 1D - input (100,1 ReLU), Conv1D(1x100x16x5 ReLU), maxPool, Conv1D(1x50x32x5 ReLU), maxPool, Conv1D(1x25x48x5 ReLU), maxPool, Conv1D(1x12x64x5 ReLU), maxPool, Conv1D(1x6x80x5 ReLU), flatten, dense(480), dropout(0,2), output(9,softmax)	98,566
CNN 1D - input (100,1 ReLU), Conv1D(1x100x16x7 ReLU), maxPool, Conv1D(1x50x32x7 ReLU), maxPool, Conv1D(1x25x48x7 ReLU), maxPool, Conv1D(1x12x64x7 ReLU), flatten, dense(768), dropout(0,2), output(9,softmax)	98,886
RNN (units=1, dropout=0.2, dense activation=softmax)	78,689
RNN (units=5, dropout=0.2, dense - activation=softmax)	34,009
RNN (units=20, dropout=0.2, activation=softmax)	78,56
RNN (units=256, dropout=0.2, dense - activation=softmax)	68,943
LSTM (units=1, dropout=0.2, dense - activation=softmax)	76,626
LSTM (units=5, dropout=0.2, dense - activation=softmax)	85,866
LSTM (units=20, dropout=0.2, dense - activation=softmax)	88,406
LSTM (units=256, dropout=0.2, dense - activation=softmax)	87,091
MLP (1 warstwa ukryta, 256 połączeń, aktywacja=ReLU, warstwa wyjściowa=softmax, dropout=0.2)	97,789
MLP (1 warstwa ukryta, 512 połączeń, aktywacja=ReLU, warstwa wyjściowa=softmax, dropout=0.2)	97,957
MLP (1 warstwa ukryta, 1024 neuronów na warstwę, aktywacja=ReLU, warstwa wyjściowa=softmax, dropout=0.2)	97,58
MLP (1 warstwa ukryta, 512 połączeń, aktywacja=ReLU, warstwa wyjściowa- softmax)	98,014
MLP (2 warstwy ukryte, 512 połączeń, aktywacja=ReLU, dropout=0.2, warstwa wyjściowa=softmax)	97,643
MLP (3 warstwy ukryte, 512 połączeń, aktywacja=ReLU, , dropout=0.2, warstwa wyjściowa=softmax)	97,674
DBN (n_komponentów=1)	72,18
DBN (n_komponentów=5)	83,786
DBN (n_komponentów=16)	91,171
DBN (n_komponentów=64)	90,04
2 input CNN - 2xInput (1x50x1), 2xConv1D(1x50x16x3),2xmaxPool, 2xConv1D (1x25x16x3), concatenate(1x25x32x3), flatten, dense (800), output(9,softmax)	97,8
2 input CNN - 2xInput (1x50x1), 2xConv1D(1x50x16x3),2xmaxPool, 2xConv1D (1x25x32x3), 2xmaxPool, 2xConv1D(1x12x32x3), concatenate(1x12,64x3), flatten, dense (768), output(9,softmax)	97,43
2 input CNN - 2xInput (1x50x1), 2xConv1D(1x50x16x3),2xmaxPool, 2xConv1D (1x25x32), 2xmaxPool, 2xConv1D(1x12x32), 2xmaxPool, 2xConv1D(1x6x64x3), concatenate(1x6x128x3), flatten, dense (768), output(9,softmax)	96,13
CNN 1D BEST - input (100,1 ReLU), Conv1D(1x100x16x3 ReLU), maxPool, Conv1D(1x50x32x3 ReLU), maxPool, Conv1D(1x25x48x3 ReLU), maxPool, Conv1D(1x12x64x3 ReLU), maxPool, Conv1D(1x6x80x3 ReLU), flatten, dense(480), dropout(0,2), dense(200), output(9,softmax)	99,6

Różne metody klasyfikacji przynosiły różne wyniki. Po optymalizacji parametrów danych metod widać, że nawet standardowe metody uczenia ML osiągały dokładność klasyfikacji przewyższającą 95%. Spośród tych metod SVM osiągnęło najwyższą dokładność dochodzącą do 98,8% poprawności klasyfikacji danych walidacyjnych. Stosowanie sieci rekurencyjnych i sieci LSTM nie przynosiło dobrych efektów. Sieci te jak wykazano w przeglądzie literatury często stosowane są w analizie przebiegów czasowych, natomiast ze względu na swoją zdolność do zapamiętywanie będą bardziej skuteczne w przetwarzaniu przebiegów wykazujących pewne powtarzalne trendy zmian. W analizowanym zagadnieniu każdy z impulsów jest odrębnym zagadnieniem klasyfikacji o dużej zmienności cech. Z tego względu klasyfikację wyładowań można porównać np. do wieloklasowego rozpoznawania obrazów. Z tego względu klasyczne w tego typu zagadnieniach sieci CNN osiągały wysokie dokładności. Zarówno w wypadku metod ML jak i sieci neuronowych widać

także, że zwiększanie złożoności rozwiązania (jak np. bardzo duża ilość gałęzi drzewa decyzyjnego, duża ilość drzew w lesie losowym czy duża ilość warstw lub neuronów w warstwie) od pewnego poziomu nie poprawia klasyfikacji, a czasami wręcz prowadzi do degradacji wyników.

Najlepsze wyniki osiągnięto stosując głębokie jednowymiarowe konwolucyjne sieci neuronowe. Wynikowa sieć posiadała warstwę wejściową (aktywacja ReLU), 5 warstw konwolucyjnych połączonych warstwami aktywacji ReLU. Każda z kolejnych warstw konwalucyjnych posiadała zwiększającą się ilość filtrów, odpowiednio 16, 32 48, 64, 80 (rozmiar 3, stride 1, „same" padding). Pomiędzy warstwami konwolucyjnymi zastosowano warstwy łączenia wybierające maksima (MaxPooling), w celu stopniowego redukowania wymiarowości, ilości parametrów i nakładu obliczeniowego. Ostatecznie, wyniki ostatniej warstwy konwalucyjnej były spłaszczane do postaci wektora i stanowiły wejście do gęstej warstwy wyjściowej MLP z funkcją aktywacji softmax.

Pomimo stosowania aktywacji ReLU w kolejnych warstwach konwolucyjnych, podczas uczenia sieci obserwowalny był proces tzw. overfittingu, który jest wynikiem zjawiska zanikającego gradientu. Uczona sieć tak dobrze uczyła się klasyfikować dane zbioru treningowego, że traciła jednocześnie zdolność do generalizowania, czyli umiejętność poprawnej klasyfikacji danych, których uprzednio nie przetwarzała. Przykład krzywych dokładności na zbiorze treningowym oraz walidacyjnym oraz odpowiadające im wartości funkcji straty w funkcji kolejnych epok pokazano poniżej (Rysunek 65).



Rysunek 65 Zjawisko overfittingu w klasyfikacji impulsów EDM

Aby zapobiegać przetrenowaniu w sieci zastosowano zaproponowaną przez Geoffreya E. Hintona et al. [321] technikę dropout. Podczas uczenia usuwano z warstw wewnętrznych,

sieci pojedyncze neurony. Taki sposób deregulacji zmusza uczoną sieć do uczenia w sposób bardziej zgeneralizowany.

W każdej turze uczenia się, każdy z neuronów jest usuwany lub zostawiany w sieci z określonym prawdopodobieństwem. Powoduje to dynamiczne zmiany architektury sieci w kolejnych epokach. W praktyce otrzymujemy sytuację, w której jeden zbiór danych został wykorzystany do nauczenia wielu sieci o różnych architekturach, a następnie model został przetestowany na zbiorze testowym z uśrednionymi wartościami wag. Przeprowadzono szereg testów dla różnych wartości prawdopodobieństwa dropout-u. Optymalne wyniki osiągnięto stosując 20% prawdopodobieństwa usunięcia neuronów. Pozwoliło to zmniejszenie overfittingu i poprawę dokładności klasyfikacji na danych testowych. Zastosowanie dodatkowej warstwy gęstych połączeń na wyjściu ostatniej warstwy konwolucyjnej sieci dodatkowo poprawiło dokładność klasyfikacji.

Ponieważ w procesie uczenia obserwowalne były wahania wartości dokładności dla danych walidacyjnych zastosowano algorytm zapamiętywania wartości wag w każdej iteracji uczenia oraz odpowiadających im wartość dokładności. Po zakończeniu uczenia wszystkich epok wczytywano model osiągający najlepsze wyniki. Model ten stosowany był w dalszych badaniach.

Tak jak opisano wcześniej w procesie uczenia w celu określenia wydajności modelu stosowano podejście k-krotnego sprawdzianu krzyżowego. Wyniki k-krotnej kroswalidacji potwierdzają skuteczność modelu wobec nieznanych danych. Zbudowany model stanowi dobry kompromis pomiędzy obciążeniem, a zbyt dużą wariancją (przetrenowaniem) w wyniku zbyt dużej złożoności. Osiągnięto dokładność sprawdzianu krzyżowego na poziomi 99,54%.

Tabela 10 Wyniki sprawdzianu k-krotnej kroswalidacji

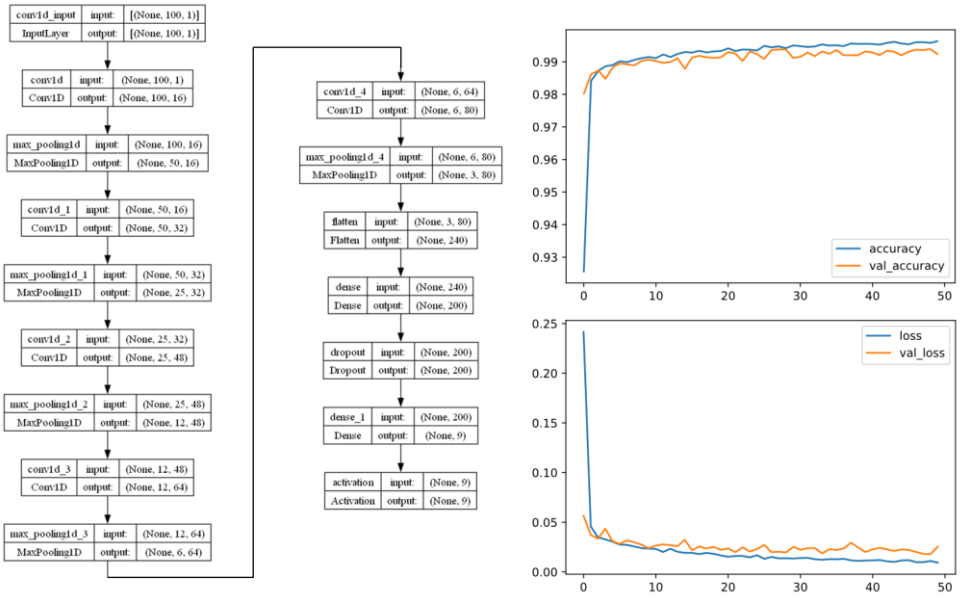
	Podzbiór 1	Podzbiór 2	Podzbiór 3	Podzbiór 4	Podzbiór 5	Podzbiór 6	Podzbiór 7	Podzbiór 8	Podzbiór 9	Podzbiór 10
Dokładność [%]	99,1614	99,4497	99,3711	99,5283	99,7379	99,4497	99,7117	99,6069	99,4497	99,8952
						Dokładność sprawdzianu [%]			99,5362 ± 0,002	

Po określeniu skuteczności sprawdzianem kroswalidacyjnym, model został ponownie poddany uczeniu z zastosowaniem całego zestawu danych uczących i sprawdzono jego skuteczność z wykorzystaniem niezależnego zestawu danych testowych. Dla tak wytrenowanego modelu określono metryki skuteczności zestawione w tabeli 11.

Tabela 11 Wyniki metryk skuteczności dla klasyfikatora impulsów

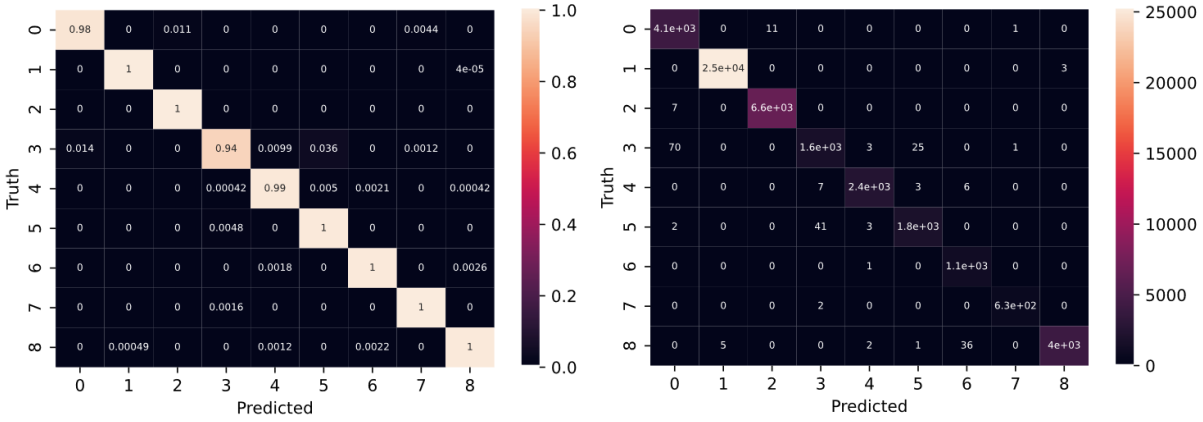
Dokładność (Accuracy)	Pełność (Recall Score)	Precyzja (Precision)	Wynik F1 (F1 score)
99,61%	99,60%	99,61%	99,60%

Poniższy rysunek 66 przedstawia budowę sieci konwolucyjnej oraz wykresy dokładności uczenia i wartości funkcji straty dla danych uczących i walidujących w kolejnych epokach procesu uczenia.



Rysunek 66 Wynikowy kształt optymalnej sieci klasyfikacji impulsów oraz wykresy dokładności uczenia oraz wartości funkcji straty dla danych uczących i walidujących w kolejnych epokach uczenia

Na poniższym rysunku 67 przedstawiono macierze błędu (po lewej wartości znormalizowane, po prawej ilości impulsów) dla przykładowego drążenia nie biorącego udziału w procesie uczenia sieci.



Rysunek 67 Macierze błędu klasyfikacji impulsów przykładowego pełnego drążenia (dane walidacyjne)

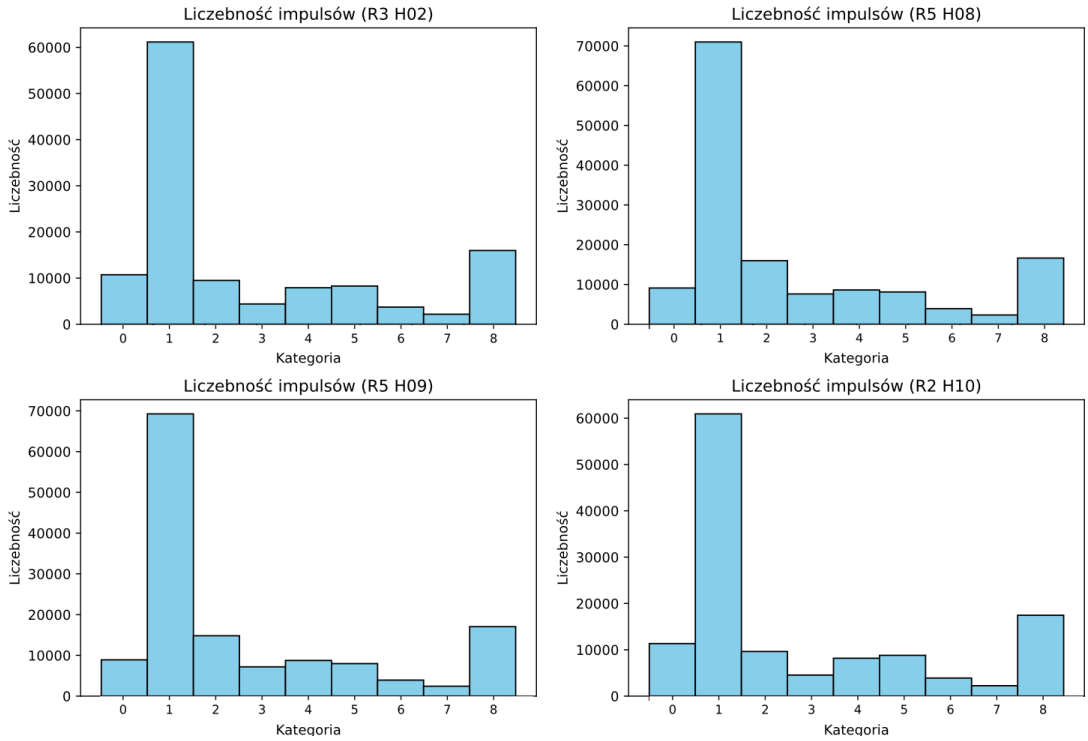
Ostateczna poprawności klasyfikacji na poziomie przekraczającym 99% jest wynikiem bardzo zadawalającym. Błędne klasyfikacje niedużej liczby impulsów z dużą dozą prawdopodobieństwa wynikają bardziej z samego charakteru analizowanych impulsów niż z niedoskonałości zastosowanej sieci neuronowej. Przegląd przebiegów błędnie sklasyfikowanych impulsów potwierdził, że miały one przebiegi, które nawet ekspertowi nie pozwalały na jednoznaczne przypisanie do tylko jednej kategorii. Błędnie sklasyfikowane impulsy klasy 3 o najmniejszej dokładności klasyfikacji (patrzy Rysunek 67), wykazywały bardzo duże podobieństwo do innych klas.

Analogiczny proces uczenia został powtórzony dla innych zestawów drążeń testowych prowadzonych z różnymi zestawami parametrów. We wszystkich przypadkach osiągnięto bardzo zbliżone wyniki dokładności klasyfikacji z niewielkimi zmianami struktury optymalnej sieci. Inne zestawy parametrów posiadały krótsze cykle EDM, przez co sieci miały mniej danych do przetworzenia co generalnie prowadziło do zmniejszenia to złożoności optymalnego klasyfikatora.

5.1.4. Analiza zmienności charakteru impulsów FHD

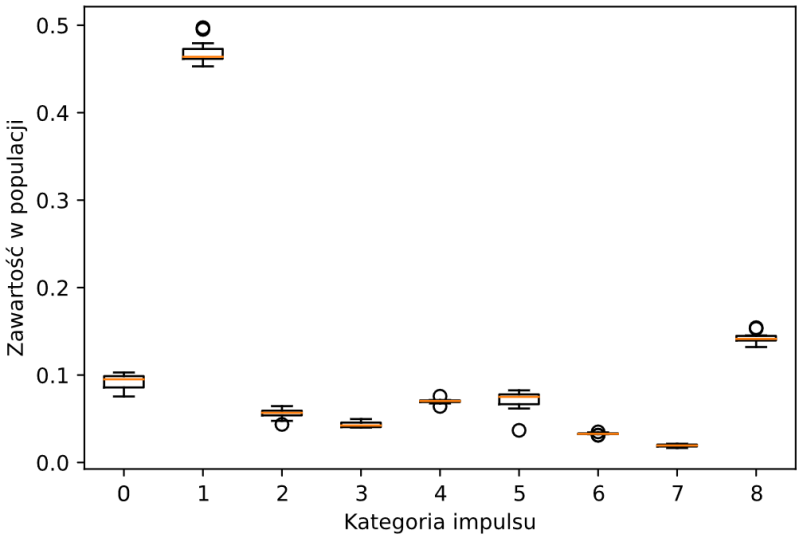
Po wytrenowaniu, optymalne sieci klasyfikacyjne zostały zastosowane do podziału impulsów na kategorie we wszystkich przebiegach próbnych. W wyniku osiągnięto przypisanie etykiety odpowiadającej danej kategorii impulsu dla każdego kolejnego okresu EDM we wszystkich zestawach danych. Pozwoliło to na bardziej szczegółową analizę procesu.

Analizując dane z przebiegów można zaobserwować rozkład liczebności typów impulsów w różnych drążeniach. Na rysunku 68 pokazano przykładowe histogramy liczebności impulsów dla wybranych drążeń z użyciem parametrów optymalnych (zestaw parametrów #1). Analizie poddawane były tylko dane zbierane w czasie drążenia, przełomu i chwilę po całkowitym przebiciu.



Rysunek 68 Liczebność różnych typów impulsów w przykładowych drażeniach testowych (parametry #1)

Powyższe histogramy są dobrą reprezentacją zależności powtarzających się we wszystkich drażeniach testowych z parametrami optymalnymi. Dane ze wszystkich drażeń testowych zostały przeanalizowane w podobny sposób. Zbiorcze wyniki analizy statystycznej wszystkich drażeń wykonanych z parametrami optymalnymi przedstawiono poniżej.



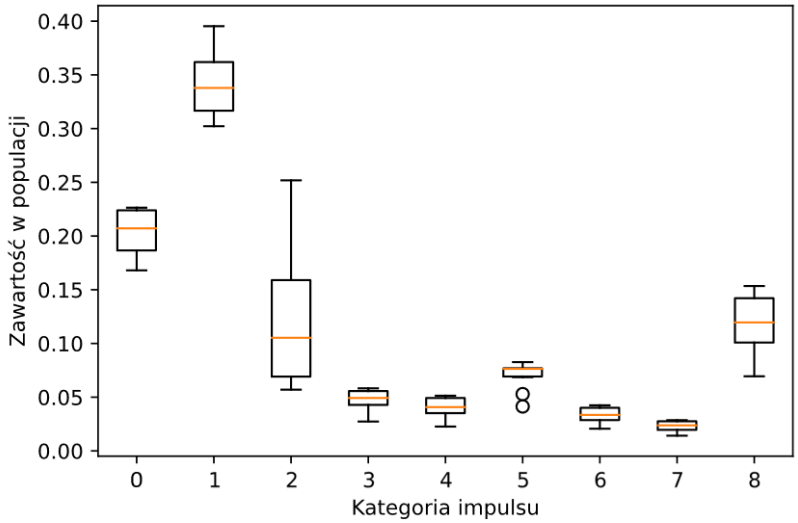
Rysunek 69 Wykres pudełkowy zawartości klas impulsów w całych drażeniach dla testów z parametrami #1

Tabela 12 Statystyki rozkładu klas impulsów w różnych drążeniach (parametry #1)

Rozkład impulsów dla parametrów #1	Kategoria 0	Kategoria 1	Kategoria 2	Kategoria 3	Kategoria 4	Kategoria 5	Kategoria 6	Kategoria 7	Kategoria 8
Minimum [%]	7.5511%	45.2989%	4.3556%	3.9944%	6.4133%	3.6778%	3.1089%	1.6600%	13.2033%
Maksimum [%]	10.2933%	49.7222%	6.4456%	4.9667%	7.5811%	8.2500%	3.5011%	2.1411%	15.4200%
Zakres [%]	2.7422%	4.4233%	2.0900%	0.9722%	1.1678%	4.5722%	0.3922%	0.4811%	2.2167%
Średnia [%]	8.9222%	47.5106%	5.4006%	4.4806%	6.9972%	5.9639%	3.3050%	1.9006%	14.3117%
Wariancja [%]	0.0073%	0.0197%	0.0033%	0.0010%	0.0007%	0.0154%	0.0001%	0.0002%	0.0038%
Odchylenie standardowe [%]	0.8535%	1.4038%	0.5742%	0.3159%	0.2714%	1.2399%	0.1113%	0.1505%	0.6126%
Skosność	-0.9076	-0.3217	-0.2044	-0.9281	0.9865	2.2169	-0.4305	-1.0312	-0.2671
Kurtoza	-0.6240	1.0351	-0.7532	0.5925	-0.1405	-1.6865	0.1618	-0.1832	0.4494

Jak widać, najwięcej okresów EDM odpowiada przebiegom otwartym (kategorie 1 i 8). Oznacza to, że bardzo często, pomimo ciągłego, cyklicznego zachodzenia czasu T_{on} wyładowanie iskrowe wcale nie występuje. Sytuacja taka może wynikać między innymi z dużego obciążenia cyklu dla zestawu parametrów #1. Nawet wyładowania o mniejszej częstotliwości mogą intensywnie usuwać materiał, co powoduje tworzenie większej szczeliny i wydłuża czas potrzebny na jej zmniejszenie i wygenerowanie impulsu. Dalej możemy zaobserwować podobne ilości zachodzących wyładowań normalnych różnych kategorii (kategorie od 3 do 7) i wyładowania łukowe (kategoria 0). W większości przeprowadzonych drążeń z parametrami optymalnymi (#1) nadal obserwowano pewną, nigdy nie zerową, ilość wyładowań zwarciowych (kategoria 2). Relatywnie nieduża ilość zwarć to sytuacja pożądana, gdyż jak opisywano w rozdziale 3.2 impulsy zwarciowe nie przyczyniają się do usuwania materiału i dodatkowo pogarszają jakość powierzchni otworu. Dla parametrów optymalnych, pomimo obserwowanej dużej, krótkookresowej zmienności generacji impulsów, zachodzi jednak duża powtarzalność średniej statystycznych ilości wystąpień danej klasy impulsów w czasie całych wierceń. Drążenia prowadzone takimi parametrami cechują się dzięki temu dużą powtarzalnością i stabilnością co jest wymogiem dla osiągania dobrych wyników produkcyjnych. Obserwacje dużej ilości wierceń potwierdzają wnioski o rozkładzie impulsów podczas drążenia wyciągnięte z analizy wykresów sylwetki (rozdział 5.1.2).

Podobną analizę przeprowadzono dla drążeń stosujących drugi zestaw parametrów (parametry agresywne). Jak opisywano w rozdziale 4.5 wstępna analiza impulsów wykazywała podobny charakter poszczególnych wyładowań, natomiast z obserwowaną różnicą w częstotliwości występowania impulsów. Analiza z wykorzystaniem sieci klasyfikacyjnych potwierdza te obserwacje. Zmiana parametrów procesu w istotny sposób zmieniła charakter częstotliwości występowania poszczególnych grup wyładowań. Wyniki analizy wszystkich drążeń testowych z parametrami #2 przedstawiono poniżej.



Rysunek 70 Wykres pudełkowy zawartości klas impulsów w całych drążeniach dla drążeń z parametrami #2

Tabela 13 Statystyki rozkładu klas impulsów w różnych drążeniach (parametry #2)

Rozkład impulsów dla parametrów #2	Kategoria 0	Kategoria 1	Kategoria 2	Kategoria 3	Kategoria 4	Kategoria 5	Kategoria 6	Kategoria 7	Kategoria 8
Minimum [%]	16,8067%	30,2211%	5,7011%	2,7344%	2,2656%	4,1367%	2,0656%	1,4156%	6,9444%
Maksimum [%]	22,6356%	39,5344%	25,1822%	5,8244%	5,1378%	8,2644%	4,2433%	2,8600%	15,3467%
Zakres [%]	5,8289%	9,3133%	19,4811%	3,0900%	2,8722%	4,1278%	2,1778%	1,4444%	8,4022%
Średnia [%]	19,7211%	34,8778%	15,4417%	4,2794%	3,7017%	6,2006%	3,1544%	2,1378%	11,1456%
Wariancja [%]	0,0444%	0,0931%	0,4162%	0,0105%	0,0099%	0,0152%	0,0059%	0,0027%	0,0852%
Odchylenie standardowe [%]	2,1065%	3,0508%	6,4512%	1,0247%	0,9929%	1,2309%	0,7691%	0,5236%	2,9192%
Skosność	-1,5028	-1,0716	-0,8740	-0,7672	-1,0689	0,5131	-1,2314	-1,3355	-1,2681
Kurtoza	-0,2818	0,4274	0,6466	-0,7151	-0,5014	-1,3566	-0,3901	-0,4210	-0,3378

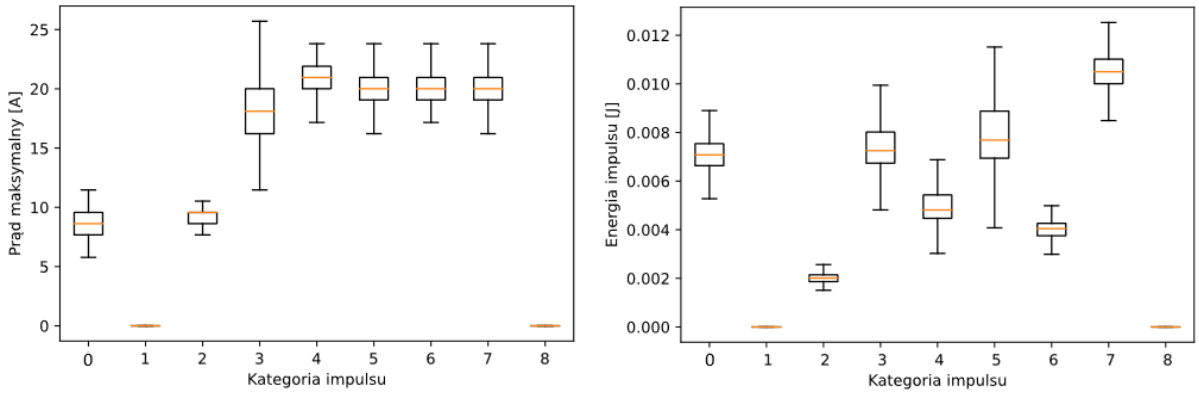
Analizując drążenia stosujące parametry agresywne (zestaw parametrów #2), widzimy różnice w rozkładzie ilości impulsów. Zauważyć można istotne zwiększenie ilości impulsów różnych klas w których dochodziło do generacji iskry, kosztem zmniejszenia częstości wystąpień obwodów otwartych. Najczęściej występującą grupą impulsów dla parametrów agresywnych były wylądowania łukowe oraz zwarcia. Wynikało to głównie ze zmniejszenia napięcia szczeliny oraz zwiększenia częstotliwości okresu EDM. Zmniejszenie napięcia szczeliny realnie zmniejsza odległość elektrody od detalu obrabianego w trakcie drążenia. Mniejsza szczelina, połączona z większą częstotliwością sygnału sterującego generatora EDM i mniejszym obciążeniem cyklu prowadzi do zwiększonej ilości impulsów. Częste impulsy oraz zmniejszenie szczeliny EDM nie pozwalają także na dokładne płukanie. Najprawdopodobniej ta sytuacja jest bezpośrednim powodem tak częstego występowania wyładowań łukowych i może powodować występowanie tzw. „miękkich" zwarć. Duża ilość wyładowań łukowych oraz ogólne zwiększenie częstotliwości zachodzenia impulsów, pomimo mniejszego obciążenia cyklu, nadal pozwala na utrzymanie dość dobrego tempa obróbki. Niestety taki przebieg procesu, szczególnie biorąc pod uwagę zwiększoną ilość zwarć, powoduje także uzyskiwanie

złej jakości powierzchni i zmiany struktury materiału w okolicach otworu, co pogarsza jego właściwości. Pogorszeniu ulega także ogólna stabilność procesu. Z analizy statystycznej rozkładu impulsów widać zdecydowanie mniejszą stabilność procesu. Obserwujemy dużą wariancję i odchylenie standardowe generacji części z grup impulsów. Jest to sytuacja niepożądana, ponieważ może przekładać się na zmniejszenie powtarzalności wyników obróbki w warunkach produkcyjnych.

Opracowane narzędzia analizy impulsów dają nam możliwość dogłębnego analizowania przebiegu procesu. Jako przykład rozszerzonego charakteryzowania procesu poniżej przedstawiono analizę rozkładu prądów maksymalnych oraz rozkład energii pojedynczego impulsu występujących w danych kategoriach impulsów. Energia w obliczeniach określana była jako całka oznaczona z mocy chwilowej pobranej przez układ:

$$w(t_1, t_2) = \int_{t_1}^{t_2} p(t)dt \tag{29}$$

Dla każdego z impulsów t_1 i t_2 odpowiadały czasom rozpoczęcia i zakończenia okresu T cyklu EDM.

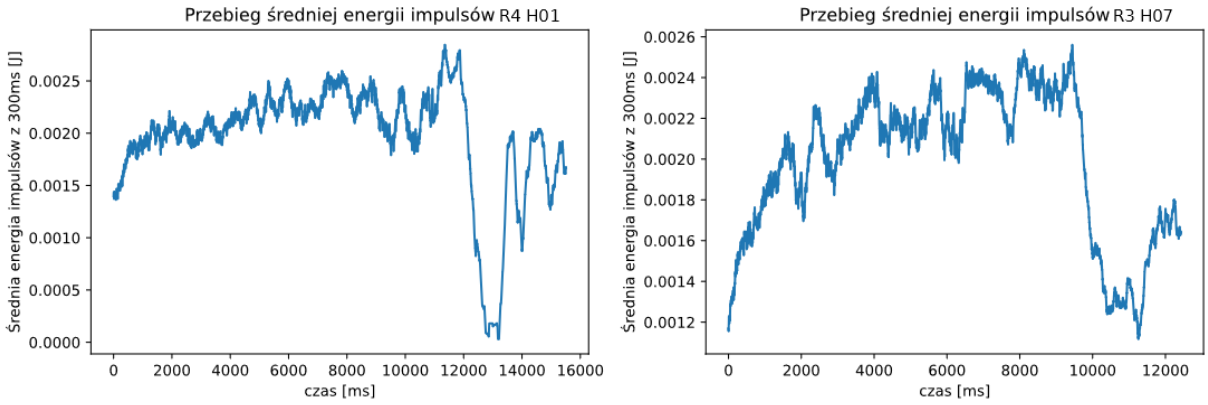


Rysunek 71 Rozkład prądów maksymalnych oraz rozkład energii pojedynczego impulsu występujących w danych kategoriach impulsów

Jak można się było spodziewać kategorie 1 i 8 odpowiedzialne za obwody otwarte mają zerowe prądy i energie. Zwarcia (kategoria 2) i wyładowania łukowe (kategoria 0) mają zbliżone prądy maksymalne, natomiast wyładowania łukowe posiadają znacznie wyższą energię ze względu na wyższy poziom napięcia w czasie ich trwania. Wyładowania normalne (kategorie 3-7) wykazują zbliżony poziom prądów maksymalnych. Wyjątkiem jest kategoria 3 gdzie rozrzut prądów jest większy. Wynika to z faktu, że te impulsy zachodziły często

przy narastającym napięciu co przekładało się na różny poziom maksymalnej szpilki prądowej impulsu. Energie impulsów są różne w zależności od typy przebiegu wyładowania. Impulsy normalne, w których występowały pojedyncze impulsy prądowe dużej wartości (kategorie 4 i 6) mają najmniejsze energie spośród wyładowań normalnych. Impulsy podwójne lub impulsy z podtrzymaniem przepływu prądu po początkowym impulsie przełamania dielektrycznego mają większą energię (kategorie 3, 5 i 7). Zmienność energii impulsu w obrębie kategorii zależy od wartości maksymalnego impulsu prądowego oraz od przebiegu podtrzymania prądu. Z tego względu impulsy wykazujące podtrzymanie prądu i następujące z opóźnieniem t_d (kategorie 3 i 5) cechują się większą zmiennością. Jak się można spodziewać, najwyższe energie wykazywały impulsy kategorii 7, gdzie przełamanie następowało na początku czasu T_{on}, po którym utrzymywał się stały przepływ prądu do jego zakończenia. Statystyczny rozkład energii impulsów dalej potwierdza prawidłowość podziału impulsów na kategorie opracowaną metodą.

Można wykreślić także wykres średniej energii przypadającej na określony okres. Wizualizacja taka może pokazać, jak zmienia się generowanie impulsów w czasie trwania całego procesu. Wykres generowany był poprzez zliczanie co 1ms średniej wartości energii 3000 poprzednich okresów EDM.

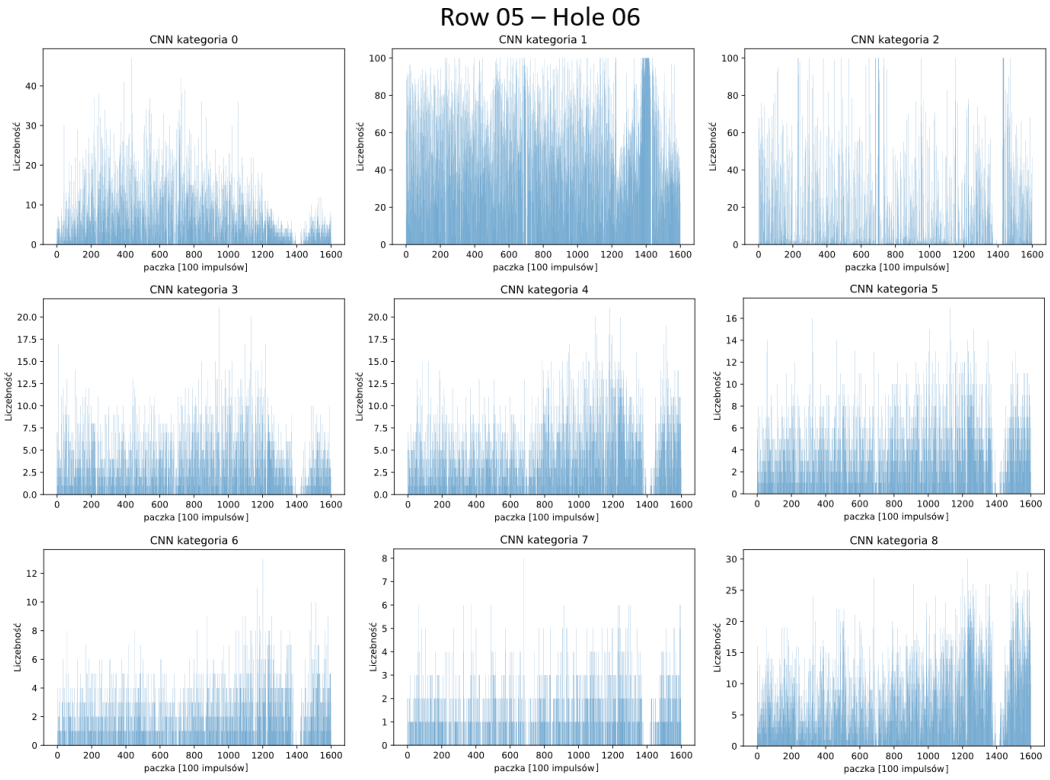


Rysunek 72 Średnia energia impulsów z okresu 300ms podczas drążenia

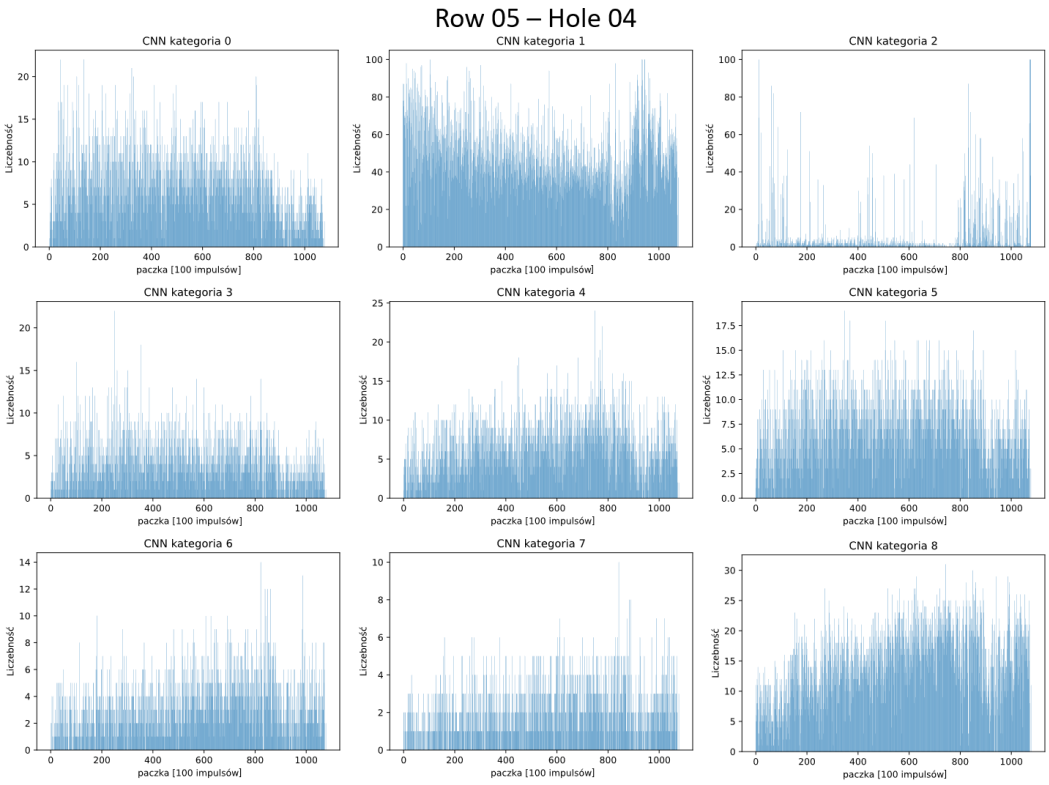
Jak pokazują powyższe wykresy (Rysunek 72), średnia energia impulsów w czasie drążenia ma pewną tendencją wzrostową skorelowaną z pogłębianiem otworu. Wynika to z faktu zwiększania powierzchni szczeliny bocznej, gdzie nachodzić mogą wtórne wylądowania. Im głębszy otwór tym więcej wyładowań bocznych co powiększa średnią energię. Wyraźnie widoczny jest także znaczny spadek średniej energii impulsów w pewnym momencie wiercenia. Zmiana ta wynika z zakończenia formowania otworu i przejazdu elektrody przez

przestrzeń pomiędzy płytkami. Taki rozkład średniej energii wyraźnie sugeruje zależność generacji impulsów od etapu procesu. Jak opisywano wcześniej w większości wierceń generacja iskier po przebiciu nie ustępuje przez co średnia energia nie spada do zera, natomiast nadal obserwowalne jest duże zmniejszenie wartości średniej po przebiciu.

Wiedząc jaki jest rozkład liczebności poszczególnych klas impulsów w drążeniach, należy sobie zadać pytanie czy ich występowanie jest skorelowane z etapem procesu. W tym celu przeprowadzono analizę częstości występowania impulsów w zależności od czasu drążenia. Dla każdego z analizowanych wierceń, cały okres drążenia został podzielony na okresy odpowiadające stu kolejnym impulsom (10ms dla parametrów #1), dalej nazywane paczkami. W każdym z okresów zsumowano ilość wystąpień impulsów każdej z kategorii. Ilości impulsów w kolejnych odcinkach czasu (paczkach) zobrazowano następnie na poniższych wykresach słupkowych. Wykresy przedstawiają dwa przykładowe wiercenia dla otworów R5 H6 i R5 H4. Otwory te zostały wybrane w celu obrazowania zależności w generacji impulsów w okolicach przełomu i przebicia dla dwóch skrajnych przypadków. W otworze H6 można było zaobserwować widoczną przerwę w generacji iskier po przebiciu, natomiast w otworze H4 iskry były generowane praktycznie przez cały czas drążenia, włącznie z etapem V (przejście elektrody przez wnękę między płytkami).

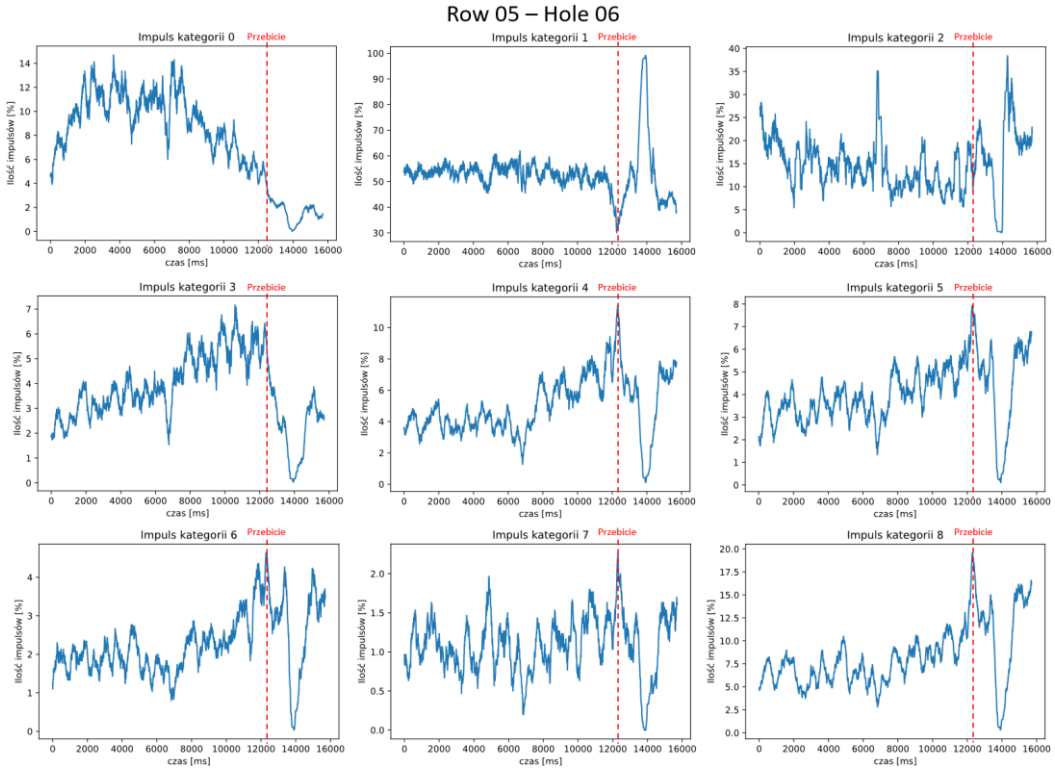


Rysunek 73 Liczebność typu impulsów w funkcji czasu wiercenia (otwór R5 H6)

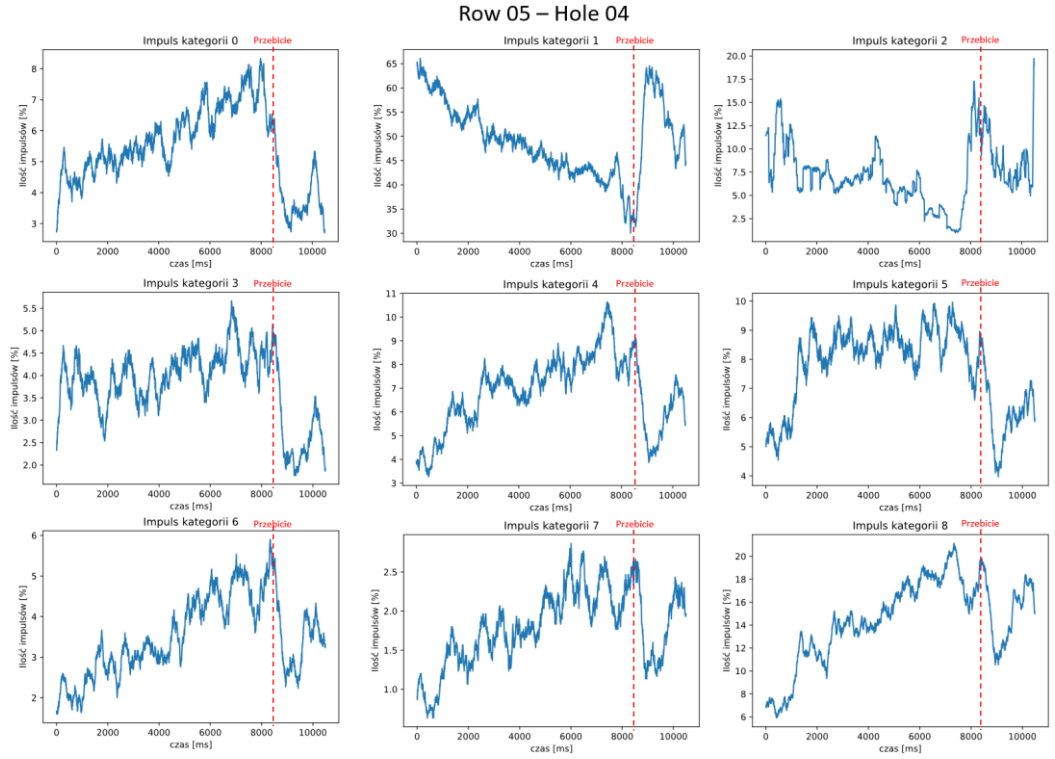


Rysunek 74 Liczebność typu impulsów w funkcji czasu wiercenia (otwór R5 H4)

Analizując powyższe wykresy (Rysunek 73 i Rysunek 74) liczebności występowania poszczególnych klas impulsów można zaobserwować ciekawe zależności. Podczas wiercenia (paczki danych od 0 do ok 1200 dla H06 i paczki od 0 do 800 dla H04) widzimy przeważającą większość obwodów otwartych, co potwierdza wyniki histogramów i wcześniejsze obserwacje szeregów czasowych. Inne klasy impulsów normalnych są w większości generowane losowo z różnymi prawdopodobieństwami. Przegląd dużej ilości analiz z różnych wierceń potwierdza zmienność rozkładu impulsów. Ilość impulsów w obrębie kolejnych 100 impulsów ma bardzo duży rozrzut. Liczebność większości klas impulsów normalnych wacha się między 0, a ok 20 wystąpieniami na 100 impulsów. Ogólna ilość zwarć (kategoria 2) w populacji jest nieduża, natomiast występują one w gęstych skupiskach sięgających nawet 100 na 100 impulsów. Potwierdza to obserwacje opisywane w przeglądzie charakteru wyładowań (rozdział 4.5). Ze względu na bardzo dużą zmienność sygnałów w obrębie 100 kolejnych impulsów, analiza trendu zmienności występowania danej klasy wyładowań w zależności od etapu procesu jest trudna. Dlatego przygotowano dodatkowe wykresy przedstawiające średnią kroczącą ilości wystąpień kategorii impulsu (Rysunek 75 i Rysunek 76). Co 1ms liczono procentową zawartość impulsów danej kategorii w 3000 poprzednich okresów EDM.



Rysunek 75 Średnia krocząca procentowej ilości impulsów (R5 H06)



Rysunek 76 Średnia krocząca procentowej ilości impulsów (R5 H04)

Na podstawie tak przygotowanych przebiegów łatwiej zaobserwować pewne zależności generacji impulsów podczas prowadzenia procesu, szczególnie w okolicach szacowanego

momentu przebicia otworu i wcześniejszego zapoczątkowania przełomu. Szacowany moment przebicia określany był na podstawie danych z maszyny (moment przebicia określony wbudowanym algorytmem analizy danych procesu) oraz poprzez wsteczną projekcję przemieszczenia elektrody wzdłuż osi Z, od momentu rozpoczęcia drążenia w płytce 2. Po przebiciu płytki 1 ze względu na zaprzestanie generacji iskier w szczelinie dolnej (przejazd elektrody przez wnękę) długość elektrody nie powinna ulegać zmianie, co pozwalało na prawidłowe określenie czasu przejazdu elektrody przez całą wnękę między płytkami, a tym samym przybliżone określenie chwili czasowej pełnego przebicia płytki 1. Obie metody określania momentu przebicia dawały zbliżone wyniki. Zaobserwować można było lekkie opóźnienie w rozpoznawaniu przebicia przez maszynę, ze względu na analizowanie zmienności trendów dla wartości uśrednianych wielkości elektrycznych, co wprowadzało przesunięcie czasowe. W pojedynczych drążeniach testowych zidentyfikowano duże opóźnia wykrycia lub całkowity brak wykrycia przebicia przez maszynę, co potwierdza ograniczenia tej metody rozpoznawania etapu procesu.

W badanych przebiegach zaobserwowano powtarzalne trendy zmienności generacji poszczególnych kategorii wyładowań. Na pierwszy rzut oka, zastawiający może być obserwowany trend zwiększania ilości generacji impulsów z czasem drążenia. Fakt ten łatwo wyjaśnić pogłębianiem otworu z czasem wiercenia. Powoduje to stałe zwiększanie powierzchni szczeliny bocznej, w której mogą występować dodatkowe impulsy zwiększające ich średnią ilość. Podczas generacji przełomu i kończenia otworu zaobserwowano spadek ilości wyładować łukowych (kategoria 0). Analizując rozkład liczebności obwodów otwartych (kategoria 1 i kategoria 8) widzimy wstępne zmniejszenie ich ilości w okolicach szacowanego momentu generacji przełomu z następującym po nim silnym zwiększeniu ich liczebności po zakończeniu tworzeniu otworu. Zmniejszenie ilości obwodów normalnych zachodzi w połączeniu ze zwiększeniem ilości wystąpień różnych klas impulsów normalnych oraz zwarć (kategorie 2-7). Częstsze generowanie impulsów wynika z opisywanych wcześniej zjawisk zachodzących w pierwszych etapach generowania przełomu. Duża ilość zanieczyszczeń dielektryka i złe warunki płukania w szczelinie bocznej oraz pogorszenie stabilizacji pozycji elektrody mogą przyczyniać się do większego prawdopodobieństwa wystąpienia wyładowań. Potwierdza to tezę stawianą w rozdziale 4.4 o zmianie średnich wartości prądu i napięcia w momencie wytworzenia przełomu. Po całkowitym uformowaniu otworu ilość wyładowań normalnych wyraźnie zmniejsza się. Średnia ilość zwarć (kategoria 2)

podczas drążenia utrzymywana jest zwykle na relatywnie stałym poziomie, natomiast w przebiegach, w których nie występuje całkowite zatrzymanie iskrzenia po przebiciu zaobserwować można zwiększenie ilości zwarć w tym okresie. Dla otworu H6 widzimy całkowite zatrzymanie generowania iskier (~13,8 sekunda) utrzymujące się do dojazdu do płytki 2 i ponownego rozpoczęcia drążenia. Nawet w otworze H4, pomimo ciągłego generowania iskier po wytworzeniu przełomu, można zaobserwować trend zwiększenia ilości obwodów otwartych (okolice 9 sekundy), który korelowany jest z zakończeniem otworu w płytce 1. Potwierdza to założenia o zmianie charakteru generacji impulsów poczynione na podstawie analizy literaturowej [64, 65, 69]. Co ciekawe, publikacje naukowe w tym temacie przedstawiają znacznie bardziej wyidealizowane przebiegi. W pracach badawczych granica przebicia jest przedstawiana jako bardzo widoczny próg zmiany generacji iskier. W przeprowadzonych badaniach obserwowano zdecydowanie mniej oczywiste granice faz procesów włącznie z ciągłą generacją iskier nawet podczas przejazdu przez wnękę między płytkami testowymi, co widać na przykładzie opisywanego otworu H4. Taki przebieg procesu znacząco utrudnia zadanie poprawnego wykrycia przebicia.

Pomimo opisywanych powyżej utrudnień, dogłębna analiza charakteru procesu, w szczególności zmienności różnych klas impulsów w zależności od etapu procesu potwierdza tezę o możliwości bazowania wykrywania przebicia na podstawie wysokoczęstotliwościowych sygnałów prądu i napięcia. Obserwowana powtarzalna zmienność wyładowań w etapie przełomu i przebicia pozwala twierdzić, że możliwe jest zbudowanie klasyfikatora, który będzie automatycznie wykrywał te zmiany i prawidłowo identyfikował etap procesu na ich podstawie.

5.2. Rozwiązanie wykrywania przebicia oparte o metody DL

Jak pokazała analiza przebiegów kategorii impulsów z poprzedniego rozdziału można przyjąć tezę, że możliwe jest zbudowanie klasyfikatora opartego o metody DL, który na podstawie danych pomiarowych będzie w stanie rozpoznać moment przebicia. Aby zbudować takie rozwiązane zastosowano dwa główne podejścia:

- Oparcie klasyfikatora o analizę większych zestawów (paczek) danych
- Oparcie klasyfikacji na testowanym wcześniej klasyfikatorze impulsów

Pierwsza z metod zakłada zbudowanie klasyfikatora, który jako wejście będzie przyjmował całe pakiety pomiarów z określonego odcinka czasu i na podstawie takich surowych danych będzie samodzielnie próbował określić moment przebicia.

Druga metoda zakładała wstępne klasyfikowanie każdego z impulsów w czasie drążenia. Dalej system może przechowywać bufor danych zawierający przydzielone etykiety klas impulsów z pewnego okresu. Bufory zawierające etykiety kategorii następnie, okresowo podawane są na wejście dodatkowej sieci neuronowej, której zadaniem jest określenie momentu przebicia.

5.2.1. Wykrywanie przebicia w oparciu o analizę zestawów danych

Podejście wykrywania przebicia oparte o analizę zestawów danych zakłada możliwość zbudowania modelu przyjmującego jako wejście całe zestawy wysokoczęstotliwościowych danych prądu i napięcia. Model następnie powinien być w stanie samodzielnie wyekstrahować z nich istotne cechy i na ich podstawie określić czy dany zestaw danych wejściowych został zebrany w czasie normalnego drążenia czy po uformowaniu całego otworu.

W procesie uczenia sieci, tak jak przy klasyfikatorach impulsów stosowano metodę wydzielania oraz k-krotny sprawdzian krzyżowy do określania skuteczności badanych modeli. W celu optymalizacji stosowano metody przeszukiwania siatki i przeszukiwania losowego. Wszelkie pokazywane metryki sprawności modeli uzyskiwane były z użyciem zbiorów danych walidacyjnych niebiorących udziału w uczeniu. Dodatkowo, spośród wszystkich wykonanych otworów losowo wydzielono zestaw 10% drążeń. Dane z tych wierceń nie brały udziału w procesie uczenia i zostały wykorzystanie do niezależnego walidowania opracowanych metod dla pełnych przebiegów procesu wiercenia.

Jako dane uczące dla sieci wykrywania przebicia na podstawie paczek danych zastosowano fragmenty przebiegów prądu i napięcia z otoczenia szacowanego momentu przebicia. Analizie poddano 500 000 próbek poprzedzających przebicie i 200 000 próbek po nim dla wszystkich otworów uczących. Dane przed przebiciem oznaczone zostały etykietą odpowiadającą normalnemu drążeniu, dane po przebiciu zostały oznaczone kolejną etykietą. Podejście takie miało na celu zrównoważenie ilości danych różnych typów w populacji uczącej. Przeprowadzając uczenie na danych z całych drążeń wprowadzilibyśmy bardzo dużą przewagę

normalnych przebiegów drążenia i proporcjonalnie bardzo małą ilość danych odpowiedzialną za identyfikację przebicia, co niekorzystnie wpływało na wyniki klasyfikacji. Zawierając w danych uczących dłuższy fragment przebiegów przed przebiciem wprowadzaliśmy odwzorowanie w danych zarówno stanu normalnego drążenia, jak i zjawisk zachodzących w momencie utworzenia przełomu.

Następnie cała baza danych dzielona była na zestawy danych o określonej długości, które dalej stanowiły dane uczące i testowe dla sieci neuronowej. Podejście takie opierało się na obserwowanej częstości występowania impulsów opisywanej w poprzednim rozdziale. Ze względu na dużą krótkookresową zmienność przebiegów, analizie poddawane były dłuższe paczki danych, które powinny z założenia lepiej oddawać prawdziwy charakter procesu w danej chwili, a co za tym idzie mogą pozwalać na identyfikację przebicia. Całość zestawów danych z kolejnych drążeń dzielona była na kolejne paczki w krokach co 2500 próbek, co odpowiada analizie procesu co 5ms.

Jako model klasyfikatora testowano róże warianty podejść ML oraz głębokich sieci neuronowych, z różnymi hiperparametrami i budowami. Bardzo istotnym wyzwaniem pracy był dobór modelu dającego jak najwyższą dokładność, na poziomie jak najbliższym 100%, przy jednoczesnej relatywnie niskiej złożoności. Tak wysoka dokładność jest wymagana ze względu na typ zadania jakim jest wykrywanie przebicia. Błędne zidentyfikowanie tego stanu podczas drążenia w etapie 3 (błąd 1 rodzaju), w warunkach produkcyjnych spowoduje zatrzymanie drążenia i utworzenie niedrożnego otworu. Otwór taki musi być poddany naprawie co jest sytuacją niedopuszczalną. Dodatkowo, dobrana złożoność modelu powinna pozwalać na pracę modelu online. Opracowane metody muszą działać w trakcie prowadzenia procesu, dając poprawne wyniki oraz bez wprowadzania dużego opóźnienia, które mogłoby doprowadzić do błędu overdrillingu (patrz rozdział 2.3.2).

W pierwszej fazie testów, podobnie jak w przypadku klasyfikacji impulsów EDM, podjęto próbę zastosowania klasycznych podejść ML do klasyfikacji przebicia. Każda z metod testowana była w różnych wariantach, poniższa tabela zestawia najlepsze osiągnięte wyniki.

Tabela 14 Wyniki wykrycia przebicia metodami ML

Metoda klasyfikacji	Dokładność na danych walidacyjnych dla napięcia+prądu [%]
Regresja logistyczna (odwrotność siły regularyzacji=1, kara=l2)	55,31
SVM (kernel=linear, C=2)	55,1
SVM (kernel=rbf, C=2)	70,54
kNN (2 sąsiadów, metryka minkowskiego, parametr=l2)	70,71
Drzewo decyzyjne (kryterium=gini, głębokość=1000)	73,14
Las losowy (kryterium=gini, ilość drzew=1000)	87,28

Jak obrazuje Tabela 14, podejście oparte o standardowe metody uczenia maszynowego nie jest wystarczające do poprawnego wykrywania przebicia. Spośród testowanych metod lasy losowe osiągały najlepsze wyniki. Niestety dokładność klasyfikacji sięgająca zaledwie 87% eliminuje to podejście. Tak duża ilość błędnych wykryć przebicia całkowicie uniemożliwiałaby prowadzenie drążeń ze względu na ciągłe zatrzymywanie procesu przed zakończeniem otworu.

Biorąc pod uwagę dobre wyniki uzyskiwane w procesie klasyfikacji impulsów, pace w tym wariancie wykrywania przebicia skoncentrowano na sieciach CNN. Ze względu na zdolność sieci konwolucyjnych do ekstrakcji różnych map cech i efektywne rozpoznawanie wzorców na różnych poziomach oraz relatywnie niedużą złożoność obliczeniową osiąganą dzięki redukcji wymiarowości pomiędzy warstwami, podejście oparte o CNN może stanowić skuteczne narzędzie rozpoznawania przebicia. Badaniom poddane zostały dwa główne podejścia: sieci 1D, oraz sieci 2D ze wstępnym przetwarzaniem sygnałów. Jak opisywano w rozdziale 0, wielu badaczy [203-209] stosowało przetwarzanie danych czasowych do formy czasowo-częstotliwościowej lub inne transformacje w celu skorzystania z dobrych zdolności klasyfikacji modeli 2D.

W badaniach sieci 2D wykorzystano ciągłą transformatę falkową jako przykład transformacji czasowo częstotliwościowej oraz zaawansowane metody wizualizacji przebiegów czasowych, takie jak Pole Kątowe Gramma GAF (ang. Gramian Angular Field) oraz Pole Przejść Markova MTF (ang. Markov Transition Field) zaproponowane przez Wanga et al. [322].

W celu stworzenia Pola Kątowego Gramma najpierw normalizowano dane wejściowe do zakresu [-1, +1]. Dla szeregu czasowego $X = \{x_1, x_1, \dots, x_n\}$ dane znormalizowane oznaczone zostają jako $\tilde{x_i}$.

Następnie każda wartość w przekształcona zostaje na współrzędne biegunowe, zgodnie ze wzorem:

$$\phi_i = \arccos \widetilde{x_i}$$
$$r_i = \frac{t_i}{N} \tag{30}$$

Gdzie $t_i \in N$ jest znacznikiem czasowym punktu danych.

Ostatecznie, metoda GAF definiuje swój własny „specjalny" iloczyn skalarny jako:

$$\langle x_1, x_2 \rangle = \cos(\phi_1 + \phi_2) \tag{31}$$

Korzystając z powyższych zależności metoda GAF buduje macierz $G = \tilde{X}^T \tilde{X}$:

$$G = \begin{bmatrix} \langle x_1, x_1 \rangle & \langle x_1, x_2 \rangle & \dots & \langle x_1, x_N \rangle \\ \langle x_2, x_1 \rangle & \langle x_2, x_2 \rangle & \dots & \langle x_2, x_N \rangle \\ \dots & \dots & \dots & \dots \\ \langle x_N, x_1 \rangle & \langle x_N, x_2 \rangle & \dots & \langle x_N, x_N \rangle \end{bmatrix}$$

Tak stworzona macierz przedstawia obraz zależności każdego punktu względem każdego innego punktu w szeregu czasowym.

Metoda MTF oparta jest o zasady modeli Markova. Metoda ta umieszcza każdą wartość szeregu czasowego w określonej ilości podprzedziałów tzn. „dyskretyzuje" każdą wartość. Każdy przedział może być określany jako „stan" zgodnie z terminologią modelu Markowa.

Następnie tworzona jest macierz przejść stanów. Wartości komórek macierzy liczone są jako:

$$A_{ij} = P((s_t = j | s_{t-1} = i) \tag{32}$$

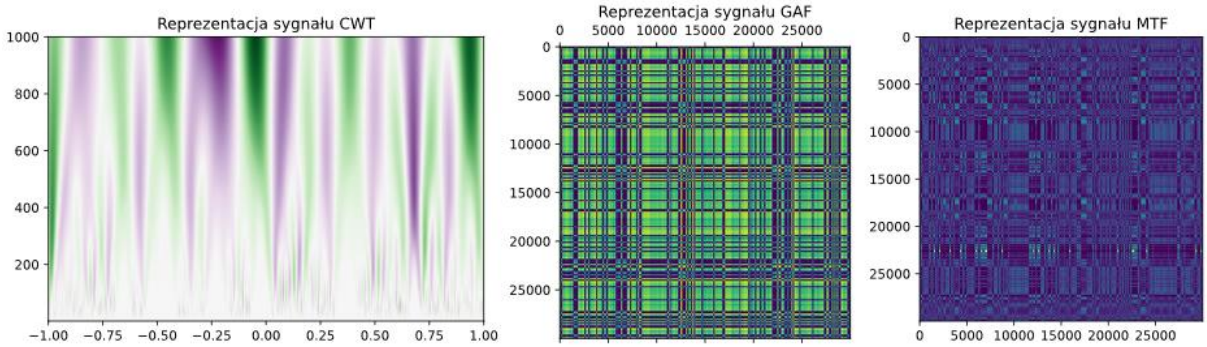
Gdzie A_{ij} jest prawdopodobieństwem przejścia ze stanu i do stanu j.

MTF działa na podobnej zasadzie, tworzy macierz *M* o rozmiarze *NxN:*

$$M_{kl} = A_{q_k q_l} \tag{33}$$

Gdzie q_k jest podprzedziałem (stanem) dla x_k, a q_l jest podprzedziałem dla x_l. Czyli M_{kl} to prawdopodobieństwo przejścia w jednym kroku z przedziału dla x_k do przedziału dla x_l. MFT obrazuje, jak bardzo dwa dowolne punkty w szeregu czasowym są ze sobą powiązane, w odniesieniu do tego, jak często pojawiają się obok siebie.

W każdym z podejść kolejne paczki danych przetwarzane były na serię macierzy („obrazów"), które następnie stanowiły wejście do konwolucyjnych sieci klasyfikacyjnych. Przykładowe zobrazowania pojedynczego zestawu danych pokazano na poniższym rysunku 77.



Rysunek 77 Przykład danych wejściowych zobrazowanych przez CWT, GAF i MTF (pakiet 30 000 próbek)

Niestety jak pokazały badania, podejście takie okazało się niepraktyczne. Wstępne przetwarzanie danych bardzo zwiększało złożoność obliczeniową i zużycie pamięci całości systemu. W przypadku każdej z testowanych metod, pojedynczy zestaw danych przetwarzany był na bardziej złożony zestaw informacji o wyższej wymiarowości. Dla GAF i MFT, w celu uzyskania pełnego odwzorowania danych konieczne było tworzenie macierzy nxn, gdzie n to ilość próbek wejściowych. W wypadku przetwarzania bardzo dużej ilości danych z wielu otworów wymaganej do dobrej generalizacji zmienności procesu i konieczności stosowania długich pakietów danych (co zostanie szerzej opisane w kolejnych paragrafach), poprawne szkolenie sieci było utrudnione. Metody te wprowadzają też duże obciążenie obliczeniowe na etapie przygotowania danych. GAF oblicza iloczyn skalarny dla każdej kombinacji dwóch punktów z paczki danych, MTF oblicza prawdopodobieństwo jednokrokowego przejścia ze stanu do stanu dla każdej pary punktów. Dodatkowym ograniczeniem podejścia wykrywania przebicia 2D, jest konieczność budowania złożonych sieci CNN 2D w celu prawidłowego rozpoznawania znacznie większej niż w wypadku samych sygnałów 1D złożoności cech wejściowych sieci. Konieczność przetwarzania danych online wymusza stosowanie podejść wydajnych obliczeniowo. Z tego względu oraz biorąc pod uwagę

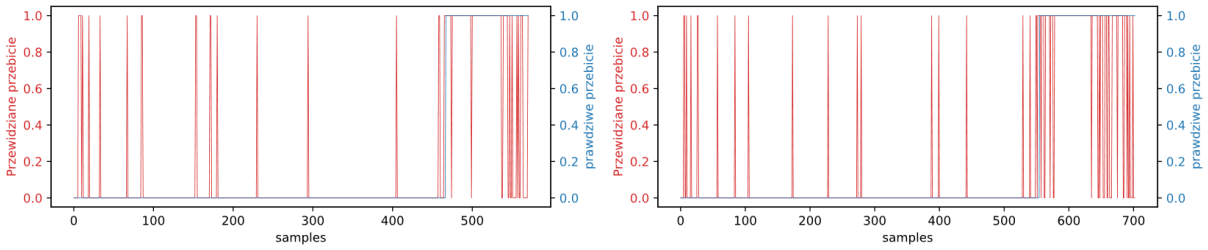
bardzo dobre wyniki opisywanych poniżej podejść CNN 1D, warianty wizualizacji danych i sieci neuronowych 2D zostały porzucone na wczesnym etapie badań.

Badania wersji podejścia CNN 1D rozpoczęto od klasyfikacji łączonych przebiegów sygnałów prądu i napięcia. Jak pokazały analizy klasyfikacji impulsów, podejście to dawało lepsze wyniki niż analiza samego napięcia lub prądu. Jednym z pierwszych zagadnień podejścia do budowy sieci było określenie wielkość zestawów danych wejściowych sieci. Przeprowadzono próby porównawcze uczenia konwolucyjnych sieci neuronowych z zastosowaniem różnych ilości danych wejściowych. Dla każdego z wariantów prowadzono optymalizację parametrów sieci w celu osiągnięcia najlepszych wyników.

Tabela 15 Osiągnięta dokładność klasyfikacji przebicia w zależności od ilości danych wejściowych i złożoności sieci

	1000 próbek	2000 próbek	5000 próbek	10000 próbek	20000 próbek	30000 próbek	40000 próbek	50000 próbek
Okno czasowe [ms]	2	4	10	20	40	60	50	100
Dokładność walidacji [%]	73,088	74,788	89,189	98,339	100	100	100	100
Ilość parametrów sieci	101058	102658	107458	115458	131458	147458	163458	179458

Jak widać w powyższej tabeli, ilość danych wejściowych sieci CNN w istotny sposób wpływała na dokładność klasyfikacji. Stosowanie krótkich bloków danych, co prawda zmniejszało złożoność sieci, natomiast nie pozwalało na osiągnięcie wymaganej dokładności klasyfikacji. Zastosowanie danych z 40ms procesu, pozwoliło na uzyskanie 100% dokładności. Niestety, jak wykazały późniejsze badania, podejście takie stwarzało problemy na etapie klasyfikacji przebiegów walidacyjnych z całych wierceń. Stosowanie danych z tak krótkiego okresu, powodowało relatywnie częste błędne określanie przebicia dla pojedynczych pakietów danych już w czasie drążenia wewnątrz płytki 1 (etap III procesu). Przykład takiego błędnego wykrywania pokazano na poniższym rysunku 78.



Rysunek 78 Błędne wykrywanie przebicia modelem o zbyt małej ilości danych wejściowych

Zjawisko to udało się znacząco ograniczyć rozszerzając ilość danych do 50 000 próbek, kosztem nieznacznego powiększenia złożoności obliczeniowej sieci. Zoptymalizowana sieć

CNN osiągnęła bardzo dobre wyniki poprawności klasyfikacji na zbiorze danych testowych. Wyniki metryk sprawności sieci przedstawiono poniżej.

Tabela 16 Wyniki metryk skuteczności dla klasyfikatora przebicia opartego o zestawy danych

Dokładność (Accuracy)	Pełność (Recall Score)	Precyzja (Precision)	Wynik F1 (F1 score)
100,00%	100,00%	100,00%	100,00%

Podjęto także próbę optymalizacji sieci do prowadzenia wykrywania przebicia z pomocą samego przebiegu napięcia. Pogłębiając sieć neuronową udało się uzyskać wyniki porównywalne do wykrywania z użyciem zarówno danych o prądzie jak i napięciu. Pomimo pogłębienia sieci (dodatkowa warstwa konwolucyjna) ogólna złożoność sieci stosującej samo napięcie była mniejsza, ze względu na mniejszą ilość danych wejściowych. Dodatkową zaletą stosowania samego pomiaru napięcia jest mniej kosztowne wdrożenie pomiarów tej wartości w warunkach produkcyjnych.

Na różnych wariantach sieci zoptymalizowanych do rozpoznawania różnej długości paczek sygnałów napięcia przeprowadzono testy wydajnościowe. Celem testów było określenie złożoności obliczeniowej poszczególnych wariantów sieci i określenie szacowanego czasu potrzebnego do inferencji podczas prowadzenia drążenia. Sprawdzano wydajność sieci z użyciem przetwarzania sekwencyjnego układu mikroprocesorowego CPU (ang. central processing unit) oraz wspomaganie przetwarzaniem równoległym z użyciem procesora graficznego GPU (ang. graphical processing unit). Test polegał na przeprowadzeniu 20 inferencji na różnych zestawach danych i określeniu średniego czasu i odchylenia standardowego przypadającego na przetworzenie jednej paczki danych. Specyfikacje sprzętu użytego w testach zamieszczono poniżej.

CPU: AMD Ryzen 6900HS (3,2 GHz, 8 rdzeni, 16 wątków)

GPU: NVIDIA RTX 2000 Ada (architektura - Ada Lovelace, 3072 potoki, 96 rdzeni Tensor/AI, zegar 2000MHz, 8GB pamięci GDDR6, przepustowość pamięci 256 GB/s)

Tabela 17 zestawia wyniki testów wydajnościowych.

Tabela 17 Wyniki testów wydajności różnych wariantów klasyfikatora przebicia opartego o zestawy danych

	10 000 próbek	20 000 próbek	30 000 próbek	50 000 próbek	100 000 próbek
Okno czasowe [ms]	20	40	60	100	200
Ilość parametrów sieci	136226	140258	144226	152226	172258
Średni czas inferencji pakietu danych -CPU [ms]	0,321970386	0,535674062	0,762968988	1,181375851	2,17203125
Odchylenie standardowe - CPU [ms]	0,02265156	0,023253289	0,031650034	0,06979827	0,083473415
Średni czas inferencji pakietu danych -GPU [ms]	0,16728299	0,179681264	0,25592285	0,344103853	0,935123514
Odchylenie standardowe - GPU [ms]	0,030348068	0,057231237	0,077897395	0,010848277	0,079386521

Jak widać z tabeli, stosowanie układów GPU poprawiło wydajność. Natomiast, ze względu na relatywnie niedużą złożoność testowanych modeli, różnica w wydajności nie jest bardzo duża. Przetwarzania z użyciem CPU było średnio tylko 2 razy wolniejsze niż w wypadku przetwarzania równoległego. Wybrany model optymalny (50 000 próbek w paczce) posiadał w sumie 152 226 trenowanych parametrów i zajmował w całościowo 594.63 KB pamięci. W testach na CPU dokonywał inferencji w zaledwie 1,181 ms co jest wynikiem zadawalającym. Oznacza to, że w czasie procesu, wykrywanie przebicia może być sprawdzane wielokrotnie szybciej niż czas potrzebny na zebranie danych wejściowych sieci. Ustalając czas wykonywania na 5ms osiągniemy wykrywanie wykonywane szybciej niż zapis danych prowadzony przez maszynę. Tym samym czas reakcji budowanego algorytmu powinien być znacznie szybszy niż algorytm zaimplementowany natywnie w elektrodrążarce.

Dla optymalnego wariantu sieci CNN 1D z 50k próbkami wejściowymi w k-krotnym sprawdzianie krzyżowym osiągnięto dokładność na poziomie 100,00%.

Tabela 18 Wyniki sprawdzianu k-krotnej kroswalidacji klasyfikatora przebicia

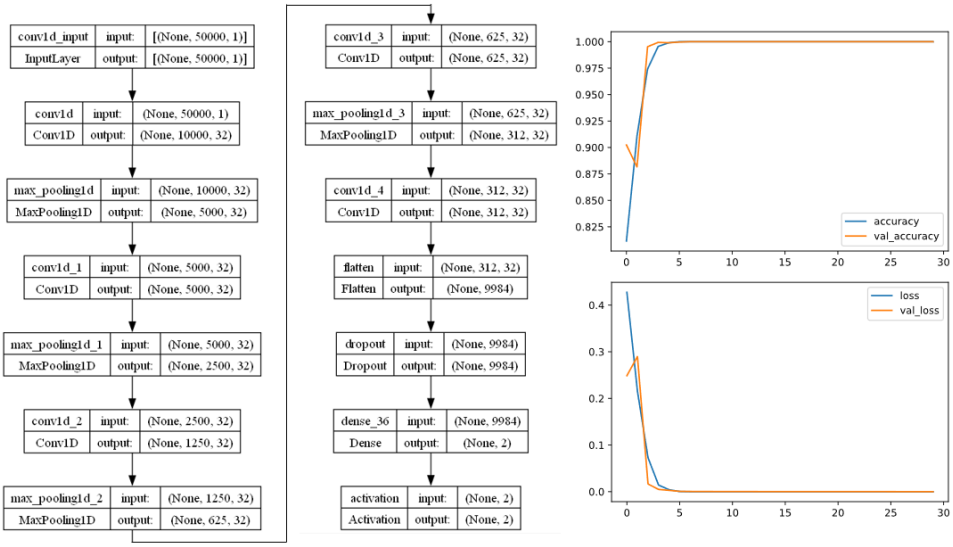
	Podzbiór 1	Podzbiór 2	Podzbiór 3	Podzbiór 4	Podzbiór 5	Podzbiór 6	Podzbiór 7	Podzbiór 8	Podzbiór 9	Podzbiór 10
Dokładność [%]	99,8556	100	100	100	100	100	100	100	100	100
						Dokładność sprawdzianu [%]			100 ± 0,000433	

Po określeniu skuteczności sprawdzianem kroswalidacyjnym, model został ponownie poddany uczeniu z zastosowaniem całego zestawu danych uczących i sprawdzono jego skuteczność z wykorzystaniem niezależnego zestawu danych testowych. Dla tak wytrenowanego modelu określono metryki skuteczności zestawione w poniższej tabeli.

Tabela 19 Wyniki metryk dokładności dla klasyfikatora przebicia stosującego paczki danych napięcia

Dokładność (Accuracy)	Pełność (Recall Score)	Precyzja (Precision)	Wynik F1 (F1 score)
100,00%	100,00%	100,00%	100,00%

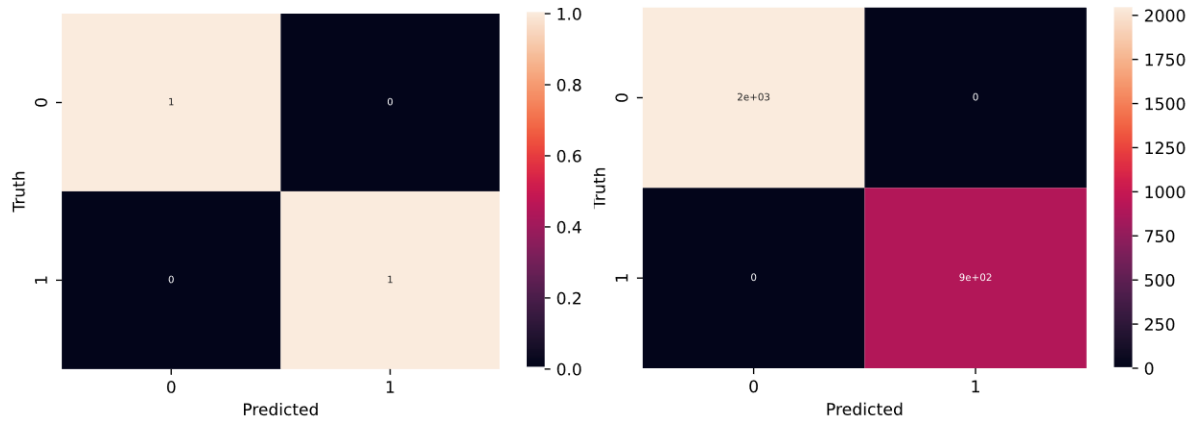
Poniższy rysunek 79 przedstawia budowę sieci CNN oraz wykresy dokładności uczenia i wartości funkcji straty dla danych uczących i walidujących w kolejnych epokach procesu.



Rysunek 79 Struktura sieci wykrywania przebicia 1D CNN oraz wykresy funkcji straty i dokładności uczenia i walidacji (50 000 próbek wejściowych)

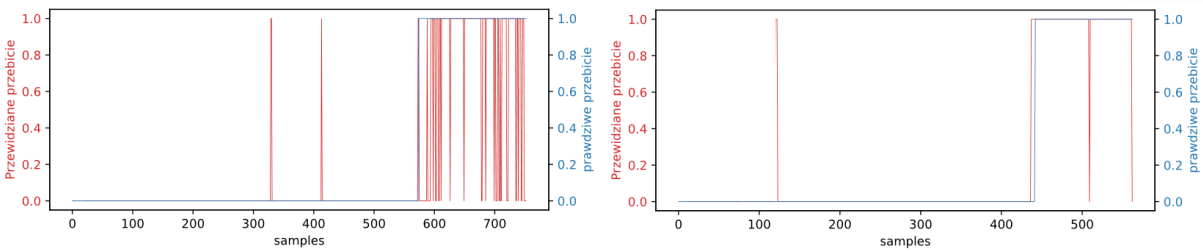
Optymalna sieć składała się z kilku sekwencyjnych warstw konwolucyjnych z następującymi po nich warstwami max poolingu z odpowiednio dobranym, zmiennym parametrem kroku (stride). Każda z warstw konwolucyjnych posiadała 32 filtry, jądro o rozmiarze 32 próbek oraz funkcję aktywacji ReLU. Wyjście sieci stanowiła warstwa gęsta z funkcją aktywacji softmax. Najlepsze wyniki w procesie uczenia osiągano z wykorzystaniem optymalizatora Adam oraz funkcji straty binarnej entropii krzyżowej. Osiągnięto wartość funkcji straty na poziomie 3,6491e-7 dla zestawu uczącego oraz 8,1995 e-5 dla zestawu danych walidacyjnych.

Na rysunku 80 przedstawiono macierze błędu uzyskane w wyniku klasyfikacji zestawu danych testowych. Podczas procesu uczenia projektowany model osiągnął 100% dokładność, nie popełniając żadnego błędu klasyfikacja na całości danych testowych.



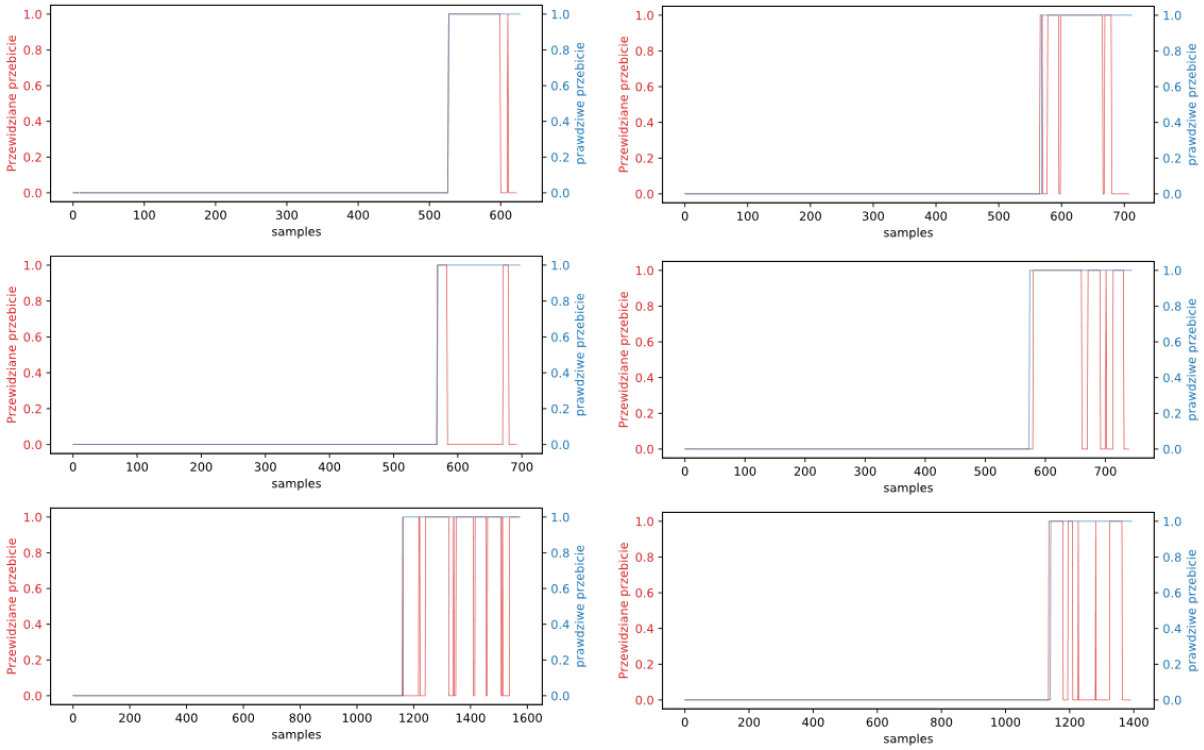
Rysunek 80 Macierze błędu wykrywania przebicia dla danych walidacyjnych (model CNN 1D – 50 000 próbek wejściowych)

Na tak przygotowanym modelu dokonano inferencji z wykorzystaniem danych walidacyjnych z całych wierceń nie biorących udziału w procesie uczenia. Całe pakiety danych wejściowych były dzielone na kolejne, sekwencyjne zestawy danych, symulujące kolejne pomiary z procesu wykonywane co 5ms, które stanowiły wejście sieci. Podejście takie pozwalało symulować inferencję w czasie drążenia, wykonywaną cyklicznie przez cały proces. Pomimo zwiększenia długości pakietów wejściowych, w pewnej liczbie drążeń walidacyjnych pojawiały się pojedyncze błędne wykrycia przebicia (Rysunek 81). Wynikało to ze zmienności procesu, nawet w prawidłowo prowadzonym drążeniu pojawiały się obszary, głównie z dużą zawartością obwodów otwartych, które mogły być błędnie interpretowane jako przebicie.



Rysunek 81 Pojedyncze błędne wykrycia przebicia modelu CNN 1D

Aby uniknąć tego zjawiska zastosowano kary nałożone na wynik inferencji. Ostateczna flaga świadcząca o przebiciu wystawiana była przez model dopiero po przekroczeniu 98% prawdopodobieństwa identyfikacji stanu przebicia oraz po zaobserwowaniu trzech kolejnych takich stanów po sobie. Pozwoliło to na prawidłowe wykrywanie przebicia, jak pokazuje Rysunek 82.



Rysunek 82 Wykrywanie przebicia ostatecznym modelem CNN 1D

Jak pokazuje rysunek, opracowane rozwiązanie pozwala na dość dobre określanie momentu przebicia. Poniższa tabela zestawia przesunięcia czasowe wykrycia w porównaniu z szacowanym momentem przebicia.

Tabela 20 Przesunięcia czasowe wykrycia przebicia z analizy pakietów danych

	Zestaw walidacyjny 1	Zestaw walidacyjny 2	Zestaw walidacyjny 3	Zestaw walidacyjny 4	Zestaw walidacyjny 5	Zestaw walidacyjny 6	Zestaw walidacyjny 7	Zestaw walidacyjny 8	Zestaw walidacyjny 9	Zestaw walidacyjny 10
Przesunięcie czasowe [ms]	72	-31	-21	-13	89	95	-7	201	105	76
						Średnie przesunięcie czasowe [ms]			56,6	

Ze względu na potencjalnie sekwencyjny charakter procesu, przetestowano także rozwiązanie kombinowane CNN-LSTM. Testowana sieć składała się z konwolucyjnych warstw wejściowych prowadzących ekstrakcję cech sygnału oraz następującej po nich warstwie sekwencyjnie ułożonych komórek LSTM. Zmieniono także podejście do procesu uczenia. Zamiast uczenia sieci losowo wybranymi pakietami danych z całej populacji, każda kolejna partia danych uczących stanowiła pełny zestaw następujących po sobie pakietów odpowiadających przebiegowi pełnego drążenia. Podejście takie powinno dawać sieci LSTM możliwość skorzystania z jej długiej pamięci krótkotrwałej w celu wykrycia długotrwałych zależności. W procesie uczenia tego wariantu sieci osiągnięto taką samą, 100% dokładność klasyfikacji, nie zaobserwowano natomiast wyraźnej poprawy wykrywania przebicia na całych

zestawach danych walidacyjnych. Wyniki wykrycia przebicia w zasadzie pokrywały się z klasyfikacją osiąganą z użyciem sieci zbudowanej w oparciu o podejście czysto konwolucyjne. Z tego względu wyeliminowano ten wariant w dalszych badaniach. Zwiększona złożoność sieci i brak poprawy wyników klasyfikacji nie dawały podstaw do prowadzenia dalszych testów rozwiązania.

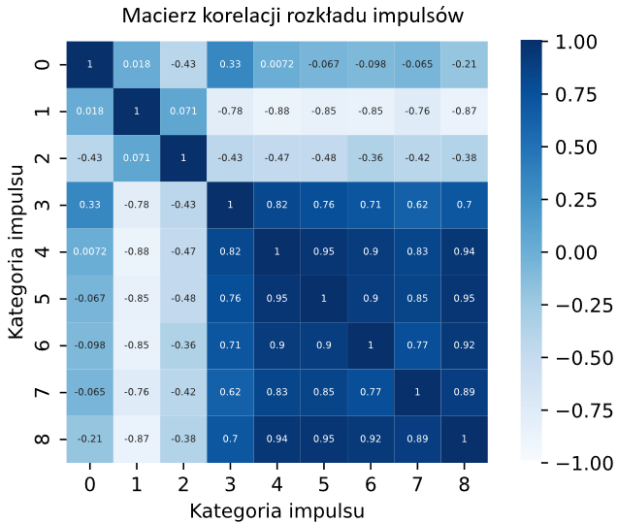
Przeprowadzone testy sugerują możliwość użycia metody jako jednego z rozwiązań wykrywania przebicia otworu w procesie FHD. Niestety, na etapie testów syntetycznych na danych z drążeń walidacyjnych zaobserwowano pewną niewielką liczbę drążeń, w których flaga przebicia występowała z przesunięciem czasowym. W ekstremalnych warunkach przesunięcie sięgało nawet 200 ms. Wynika to najprawdopodobniej z użycia szacowanych momentów przebicia otworu jako wejściowej flagi w procesie uczenia. Na etapie testów ciężko jednoznacznie określić czy przesunięcie czasowe wynika z niedoskonałości modelu CNN, czy może w wypadku wierceń, w których obserwowano przesunięcia czasowe, moment przebicia został błędnie określony już na etapie opisywania danych. W wypadku dalszego rozwoju tej metody, zasadna będzie dokładniejsza korelacja momentu wykrycia przebicia przez algorytm z faktycznym stanem na elektrodrążarce, dla dużej ilości otworów. Co istotne, w warunkach przemysłowych przesunięcie czasowe nawet na maksymalnym obserwowanym poziomie nie będzie powodowało błędu overdrillingu. W drążeniach testowych maksymalna prędkość przesuwania elektrody, niezależnie od etapu procesu, nie przekraczała 0,65mm/s. Oznacza to, że dla znacznej większości przypadków drążeń nawet przesunięcia rzędu 1 sekundy nie powodowałyby „overdrillingu". Fakt ten pozwala dopuścić rozwiązanie jako potencjalne rozwiązanie produkcyjne.

5.2.2. Wykrywanie przebicia w oparciu o klasyfikator impulsów

Kolejnym z testowanych podejść było zbudowanie narzędzia wykrywania przebicia korzystającego z opracowanej i testowanej wcześniej idei modelu klasyfikatora impulsów. Jak opisywano w rozdziale 5.1.4, w momencie przebicia można zauważyć powtarzalną zmienność w generowaniu różnych klas impulsów. Wyraźną zmianę trendu widać także na przedstawianych wykresach średniej energii impulsów. Na tej podstawie można wnioskować, że informacje o procentowych zawartościach impulsów i średnia energia (lub upraszczając, średnia moc) mogą potencjalnie stanowić dane wejściowe dodatkowej sieci neuronowej, która będzie określała na tej podstawie moment przebicia.

Pewnym utrudnieniem tego podejścia jest konieczność zapewnienia terminowości klasyfikacji impulsów EDM. Aby mieć pewność, że algorytm będzie wnioskował na podstawie pełnych danych, konieczne jest sklasyfikowanie każdego z szeregu wyładowań. Ze względu na stochastyczny charakter tworzenia impulsów i ich jego dużą krótkookresową zmienność pomijanie części impulsów, lub próbkowanie klasyfikacji ze zmniejszoną częstotliwością może prowadzić do błędnych wniosków. Z tego względu konieczne było zoptymalizowanie klasyfikatora impulsów. Zbudowane narzędzie powinno być dość proste, aby pozwalało na przetwarzanie danych co cykl EDM, co w praktyce oznacza konieczność zapewnienia czasu inferencji poniżej 100µs (parametry #1) lub 60µs (parametry #2).

Analizując zmienność średnich zawartości procentowych klas impulsów, można zauważyć dość podobne przebiegi zmienności dla impulsów normalnych (kategorie od 3 do 7). Aby lepiej zrozumieć zależności pomiędzy przebiegami generacji klas impulsów w czasie drążenia dokonano ich korelacji krzyżowej. Poniższa macierz (Rysunek 83) pokazuje uśrednione współczynniki korelacji pomiędzy przebiegami 9 klas impulsów dla wszystkich drążeń z parametrami #1. Zidentyfikować można silną dodatnią korelację zachodzącą pomiędzy różnymi klasami impulsów normalnych. Widać także silną ujemną korelację między generacją impulsów a obwodami otwartymi (kategoria 1), co jest naturalne. Zwiększenie ilości generowanych impulsów zachodzi kosztem zmniejszenia ilości obwodów otwartych.

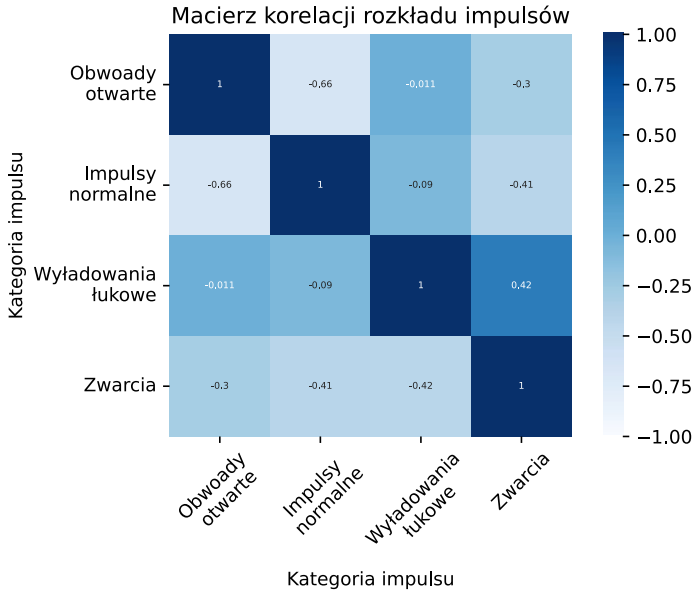


Rysunek 83 Macierz korelacji przebiegów różnych klas impulsów w czasie drążenia (parametry #1)

Ponieważ zbudowany model musi działać w czasie rzeczywistym dokonano uproszczenia konwolucyjnego klasyfikatora impulsów, w celu ograniczenia jego złożoności obliczeniowej.

Jak pokazała macierz korelacji, generacja różnych klas impulsów normalnych zachodzi z podobną zmiennością, włączając w to opisywane w rozdziale 5.1.4 zjawiska pojawiające się w okolicy przebicia. Z tego względu uproszczono analizę do wykrywania zaledwie czterech typów impulsów: obwody otwarte (wcześniej kategorie 1 i 8), zwarcia (kategoria 2), wyładowania łukowe (kategoria 0), impulsy normalne (wcześniej kategorie od 3 do 7).

Dodatkowym sygnałem, który może przynosić informacje o momencie zaistnienia przebicia mogła być opisywana w rozdziale 5.1.4 średnia energia impulsów (lub średnia moc impulsów). Głębszy przegląd zagadnienia ujawnił jednak wysoką dodatnią korelację tego czynnika z występowaniem impulsów normalnych. Obie krzywe wykazują podobne trendy zmienności w czasie drążenia i podczas kończenia otworu, co sugeruje, że będą wnosiły zbliżone, redundantne się informacje w procesie klasyfikacji. Z tego względu zmienna ta została wyeliminowana z grupy danych wejściowych modelu. Jak opisywano wcześniej, uzupełniające się informacje o wzajemnym stosunku ilości wystąpień impulsów będą dawały znacznie lepszy obraz etapu procesu.



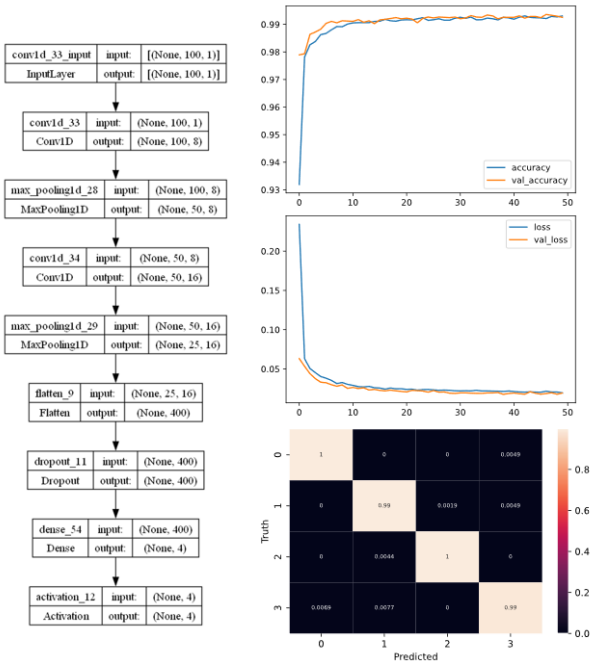
Rysunek 84 Macierz korelacji przebiegów różnych klas po redukcji do 4 kategorii

Podział impulsów na mniejszą ilość grup wykonano nadzorowanego przez użytkownika wielostopniowym procesem kategoryzacji opisywanym w rozdziale 5.1.2. Na nowo podzielonych danych powtórzono analizę korelacji przebiegów generacji impulsów. Wyniki pokazano na rysunku 84. Jak widać po zmniejszeniu ilości grup, znacznie zmniejszono

współczynniki korelacji pomiędzy nimi. Wskazywać to może, że każda z nowych kategorii niesie w sobie niezależną informację dla modelu.

Tak zmniejszona ilość kategorii impulsów powinna nadal pozwalać na określanie momentu przebicia w zależności od procentowej zawartości różnych wyładowań w różnych etapach procesu przy jednoczesnym zmniejszeniu obciążenia obliczeniowego.

Mając ponownie oznaczone dane, powtórzono proces uczenia i optymalizacji sieci neuronowej stosując takie samo podejście jak w rozdziale 5.1.3. Tym razem celem było otrzymanie znacznie uproszonego klasyfikatora, który będzie osiągał wysoką dokładność klasyfikacji.



Rysunek 85 Uproszczony klasyfikator impulsów z krzywymi straty i dokładności oraz macierz błędu dla danych walidacyjnych

Zmniejszenie ilości kategorii do 4 pozwoliło na znaczne uproszczenie klasyfikatora. Zbudowany model składał się tylko z dwóch warstw konwolucyjnych 1D z niedużą ilością filtrów. Wyjściowo klasyfikator miał jedynie 2036 uczonych parametrów i zajmował zaledwie 7,95 KB pamięci dla klasyfikatora bardziej złożonego rozpoznającego impulsy z drążeń z parametrami #1. Ostateczna formę klasyfikatora oraz wyniki uczenia i macierz błędu pokazano na rysunku 85. W tabeli 21 przedstawiono wyniki osiągniętych metryk skuteczności modelu.

Tabela 21 Wyniki metryk dokładności dla klasyfikatora impulsów z 4 kategoriami wyładowań

Dokładność (Accuracy)	Pełność (Recall Score)	Precyzja (Precision)	Wynik F1 (F1 score)
99,40%	99,36%	99,36%	99,40%

Tak mocne ograniczenie stopnia skomplikowania modelu było możliwe ze względu na bardzo wyraźne różnice pomiędzy grupami impulsów, co potwierdzają niskie wartości średniej korelacji pokazane na rysunku 84.

Aby przetestować wydajność modelu w celu sprawdzenia spełnienia krytycznego wymogu zapewnienia terminowości klasyfikacji impulsów, model został poddany serii testów wydajnościowych. Testy polegały na prowadzeniu serii 20 niezależnych inferencji różnych zestawów danych złożonych z dużej ilości okresów EDM z użyciem CPU i GPU (konfiguracja sprzętu testowego – patrz rozdział 5.2.1). Wynikiem testu było określenie średniego czasu i odchylenia standardowego przypadających na klasyfikację pojedynczego impulsu EDM. Poniższa tabela przedstawia osiągnięte wyniki.

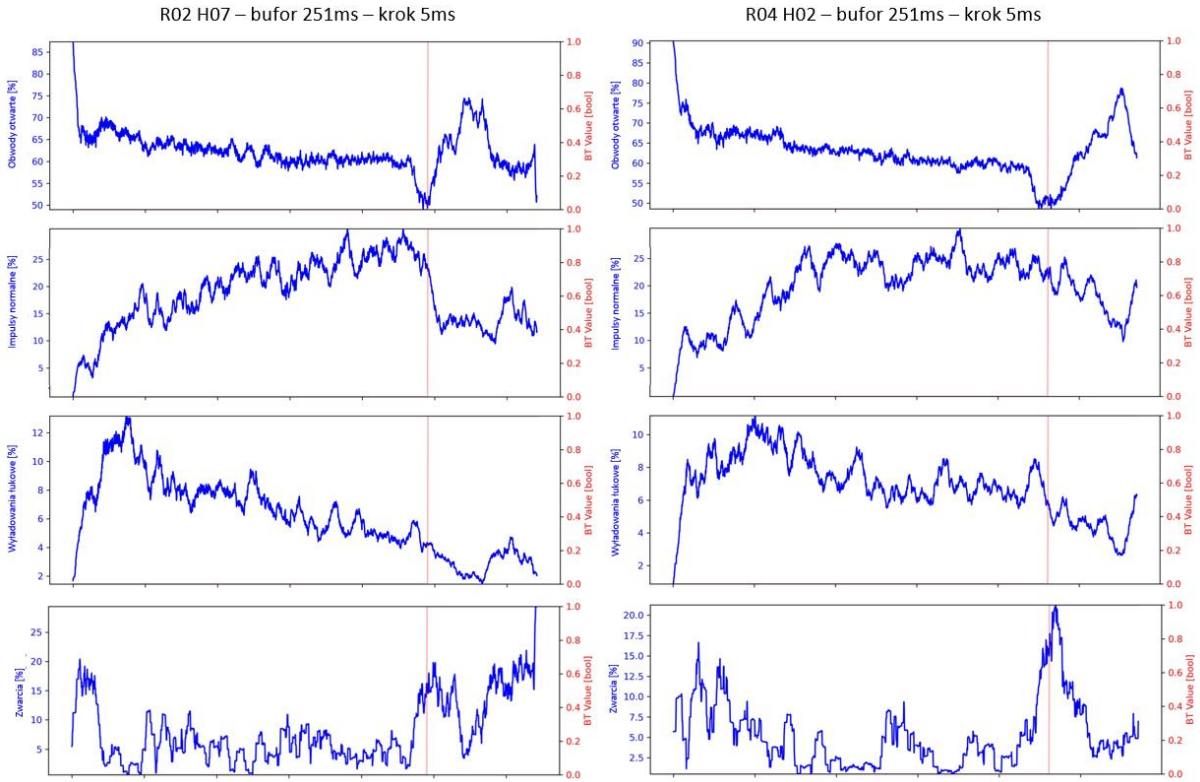
Tabela 22 Wyniki testów wydajności uproszczonej klasyfikacji impulsów na 4 kategorie

	CNN 1D 4 kategorie (parametry#1)	CNN 1D 4 kategorie (parametry#1)
Ilość parametrów sieci	2036	1196
Średni czas inferencji pakietu danych -CPU [ms]	3,33E-05	3,32E-05
Odchylenie standardowe - CPU [ms]	9,18E-07	5,22E-07
Średni czas inferencji pakietu danych -GPU [ms]	2,763E-05	2,54E-05
Odchylenie standardowe - GPU [ms]	1,37629E-06	1,26E-06

Po dokonaniu uproszczenia, zbudowane modele mogą spełniać warunki terminowości klasyfikacji impulsów na potrzeby wykrywania przebicia w czasie procesu. Jak pokazuje tabela, nawet stosując przetwarzanie z pomocą CPU maksymalny czas inferencji pojedynczego wyładowania w żadnym z testów nie przekroczył 35µs. W testowanych zestawach parametrów najkrótszy stosowany cykl EDM trwał 60 µs. Z tego względu uproszczony klasyfikator może być skutecznie stosowany.

Z pomocą tak zbudowanego klasyfikatora impulsów przeprowadzono proces przygotowania danych wejściowych dla wykrycia przebicia. Po klasyfikacji impulsów na 4 kategorie, stworzono krzywe zawartości procentowej każdej z kategorii w czasie procesu. Algorytm w odstępach 2500 próbek, co odpowiada 5ms drążenia, zliczał wystąpienia każdej kategorii impulsu z ostatnich 251 ms procesu i dzielił otrzymaną wartość przez ilość okresów EDM przypadających na ten czas. W wyniku, dla każdego drążenia testowego otrzymano 4 krzywe procentowej zawartości impulsów w czasie procesu FHD. Rysunek 86 przedstawia

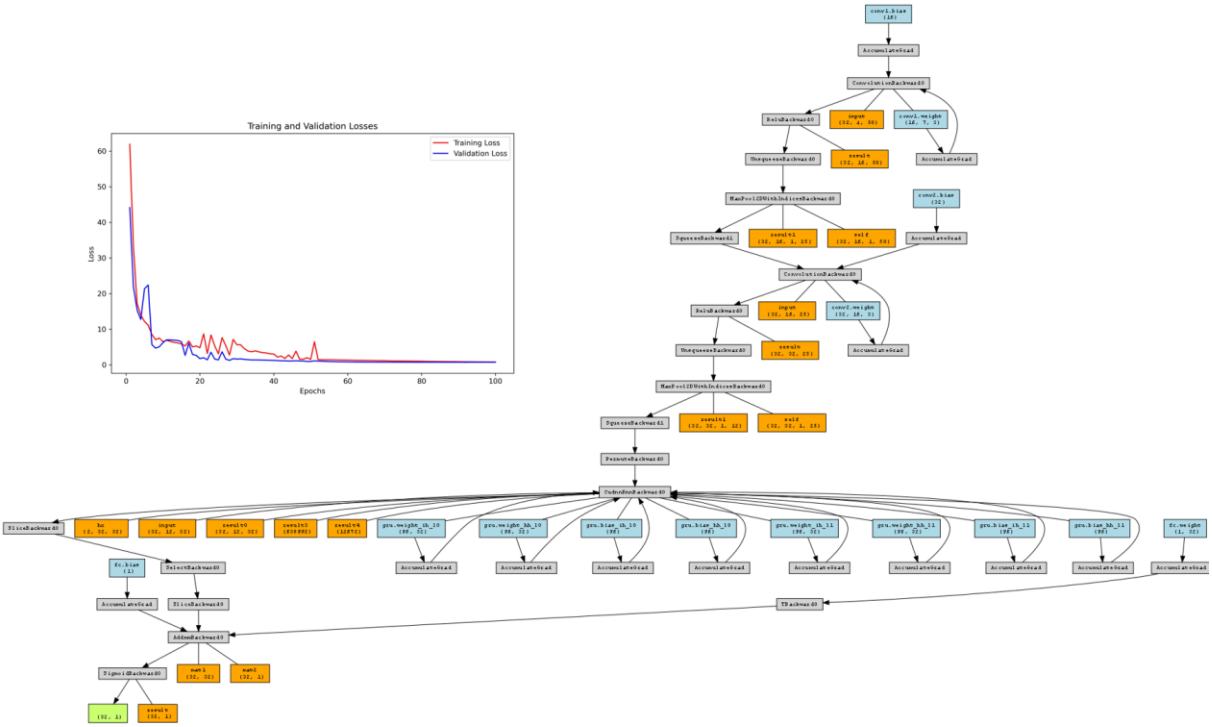
takie przykładowe krzywe z naniesioną etykietą szacowanego momentu przebicia dla dwóch przykładowych otworów.



Rysunek 86 Krzywe procentowej zawartości 4 kategorii impulsów w czasie drążenia

Tak przygotowane dane posłużyły jako dane wejściowe dla modelu wykrywania przebicia. Każdy z przebiegów zmienności zwartości impulsów stanowił w zasadzie odrębną serię czasową o nieco odmiennych przebiegach, ale podobnej zmienności wykazywanej dla różnych drążeń testowych. Z tego względu model wykrywania przebicia oparto o sieci rekurencyjne.

Zestawy danych zmienności impulsów zostały podzielone na mniejsze pakiety po 50 kolejnych wartości. Nowy pakiet tworzony był co 5 ms procesu. Dla każdego otworu testowego stworzono oddzielny zestaw sekwencyjnych pakietów, każdy z pakietów oznaczony został etykietą przed i po przebiciu. Tak przygotowane zestawy danych posłużyły jako dane uczące klasyfikatora przebicia. Sama sieć klasyfikacyjna zbudowana została z dwóch warstw konwolucyjnych prowadzących ekstrakcję cech z danych wejściowych oraz warstwę rekurencyjną z 8 komórkami. Kształt sieci oraz wykres krzywych straty pokazano na poniższym rysunku 87.



Rysunek 87 Model wykrywania przebicia oparty o GRU

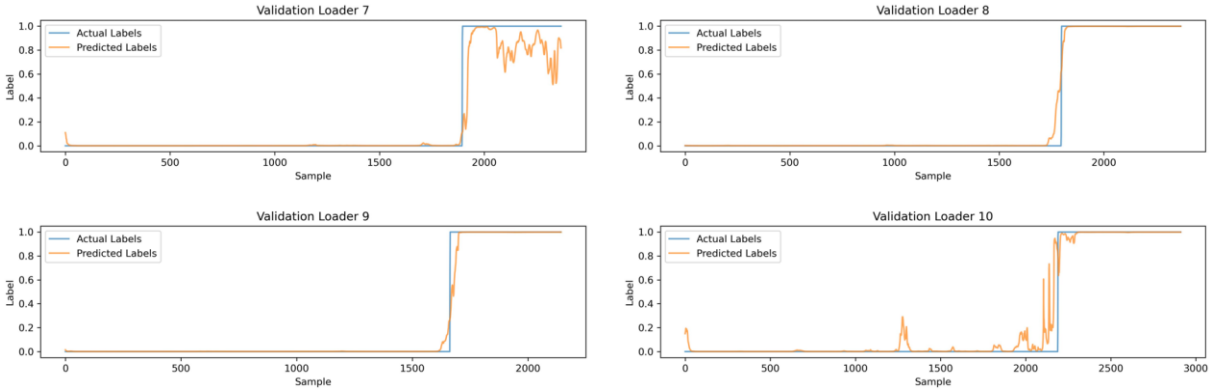
W procesie uczenia wykorzystano optymalizator Adam oraz binarną krosentropię jako funkcję straty. Przetestowano dwie formy rekurencyjnej części sieci: LSTM oraz GRU. Sieci GRU zaproponowane przez Cho et al. [323] stanowią prostszą alternatywę dla sieci LSTM. Podobnie jak LSTM, GRU może przetwarzać dane sekwencyjne, w tym szeregi czasowe.

Podstawową ideą stojącą za GRU jest wykorzystanie mechanizmów bramek do selektywnego aktualizowania ukrytego stanu sieci na każdym kroku czasowym. Mechanizmy bramek są używane do kontrolowania przepływu informacji do i z sieci. GRU ma dwa mechanizmy bramek, zwane bramką resetowania i bramką aktualizacji. Bramka resetowania decyduje, ile z poprzedniego ukrytego stanu powinno zostać zapomniane, natomiast bramka aktualizacji decyduje, ile z nowego wejścia powinno zostać użyte do aktualizacji ukrytego stanu. Wyjście z GRU jest obliczane na podstawie zaktualizowanego ukrytego stanu.

W procesie uczenia sieci stosowano podejście podobne do opisywanego wcześniej modelu CNN-LSTM (rozdział 5.2.1) opierającego się na pakietach surowych danych wejściowych. To znaczy, każda partia danych uczących sieci odpowiadała pełnemu zestawowi danych z jednego drążenia, co pozwalało sieci uczyć się sekwencyjnych zależności

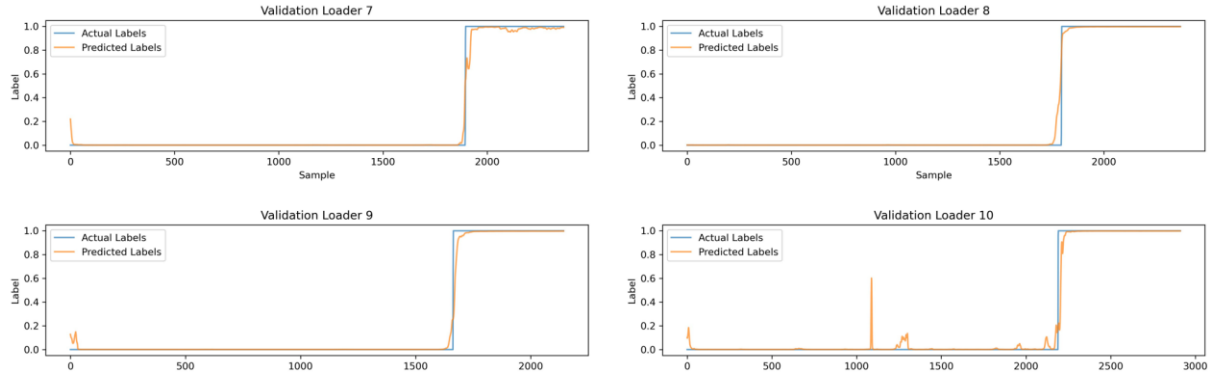
powtarzających się w kolejnych drążeniach. W procesie uczenia obserwowano czasami znaczne chwilowe wzrosty wartości funkcji straty w późniejszych epokach. Aby temu zapobiec zastosowano dynamiczną zmianę współczynnika uczenia. Co 10 epok współczynnik uczenia zmniejszany był o 20%. Podejście takie poprawiło wyniku uczenia sieci.

Po procesie uczenia zbudowane modele walidowane były poprzez przewidywanie przebicia na danych walidacyjnych z pełnych przebiegów drążenia nie biorących udziału w procesie uczenia. Kolejne paczki danych podpowiadające pomiarom z okresu 5ms były sekwencyjnie podawane na wejście modelu, który dokonywał klasyfikacji stanu.



Rysunek 88 Wykrywanie przebicia z pomocą sieci LSTM

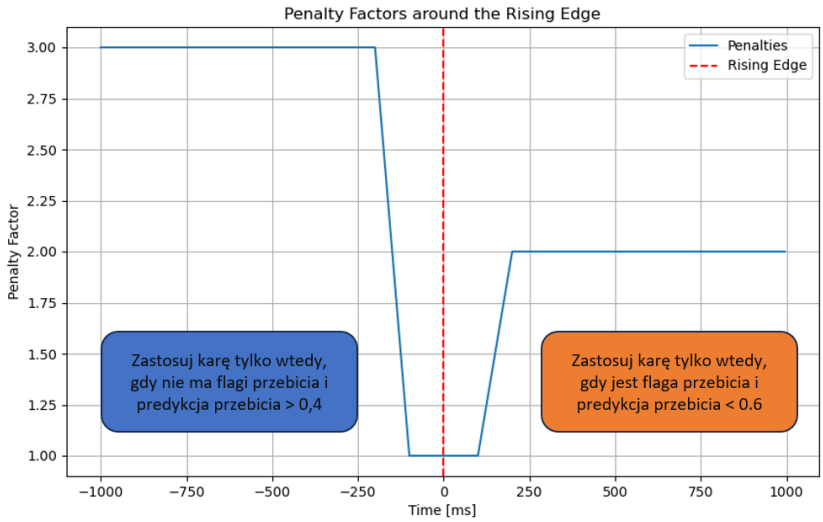
Sieć LSTM w większości przypadków dawała poprawne wyniki, natomiast w części z drążeń walidacyjnych pojawiały się pewne zawahania prawdopodobieństwa wykrycia. Moment wykrycia przebicia także, w zależności od przebiegu, określany był z przesunięciem czasowym. Przykłady takich przebiegów pokazano na rysunku 88.



Rysunek 89 Wykrywanie przebicia z pomocą sieci GRU

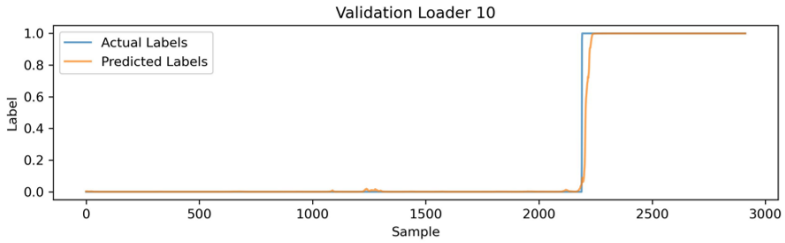
Lepsze wyniki klasyfikacji osiągano z pomocą sieci GRU. Pomimo prostszej budowy sieci tego typu, klasyfikacja w większości przypadków realizowana bała bardziej terminowo, jak pokazuje rysunek 89.

Jak widać w pewnych specyficznych przebiegach walidacyjnych, zaobserwować można było chwilowe, błędne zwiększenie prawdopodobieństwa wykrycia przebicia w czasie drążenia właściwego (przed przebiciem). Aby zniwelować to zjawisko, do procesu uczenia wprowadzono modyfikowaną funkcję straty binarnej entropii krzyżowej. Podejście modyfikuje wartości funkcji straty w otoczeniu flagi przebicia (Rysunek 90).



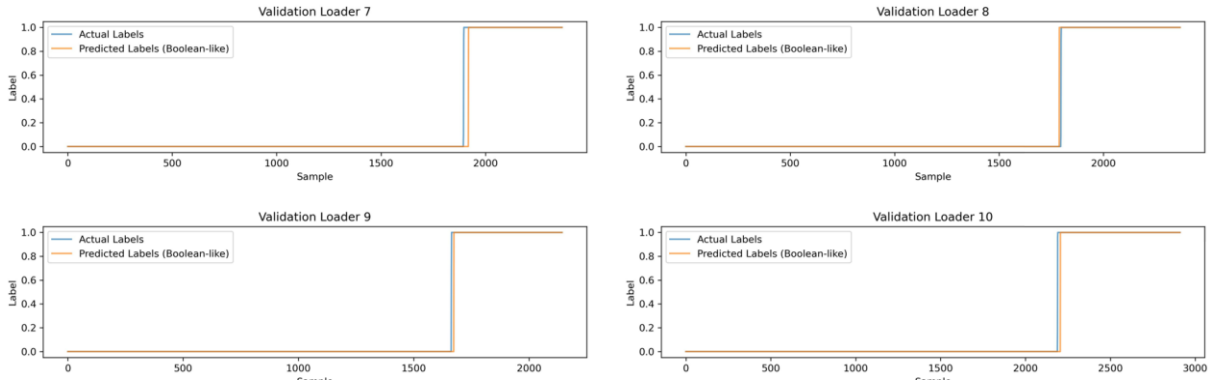
Rysunek 90 Modyfikowana funkcja straty BCE dla procesu uczenia sieci rekurencyjnych

Dzięki tak zmodyfikowanej funkcji straty silnie karane są fałszywe wczesne wykrycia przebicia. Dodatkowa kara, ale mniej surowa, nałożona jest także na fałszywe przewidywania dla danych po faktycznej fladze przebicia. Algorytm stosuje karę dla danych z czasu drążenia właściwego (przed przebiciem) jeśli prawdopodobieństwo przebicia przekroczy 40%. Dookoła faktycznego przebicia stosowana jest niezmodyfikowana funkcja BCE. Dla danych po przebiciu zastosowano kolejną karę, która jest wprowadzana, jeśli prawdopodobieństwo wykrycia przebicia spadnie poniżej 60%. Po zastosowaniu nowej funkcji straty w procesie uczenia wyeliminowano niepożądane wczesne wykrywanie przebicia (Rysunek 91).



Rysunek 91 Eliminacja błędnego, wczesnego wykrywania przebicia po zastosowaniu zmodyfikowanej funkcji straty BCE

W wypadku trenowanych sieci rekurencyjnych za każdym razem obserwowano powolne narastanie prawdopodobieństwa wykrycia przebicia w okolicach prawdziwego szacowanego momentu zakończenia drążenia otworu. Aby jednoznacznie określać moment, w którym powinien kończyć się proces wiercenia, zastosowano wartość progową równą 50% prawdopodobieństwa przebicia. Wyniki przewidywania dla końcowego modelu opartego o GRU przedstawia rysunek 92.



Rysunek 92 Wyniki wykrywania przebicia ostatecznego modelu GRU dla danych walidacyjnych

Na końcowym modelu dokonano inferencji pełnych przebiegów walidacyjnych. Osiągnięte przesunięcia czasowe pomiędzy wystawieniem flagi przebicia, a szacowanym momentem zakończenia formowania otworu przedstawiono w poniższej tabeli.

Tabela 23 Przesunięcia czasowe wykrycia przebicia z analizy przebiegu impulsów

	Zestaw walidacyjny 1	Zestaw walidacyjny 2	Zestaw walidacyjny 3	Zestaw walidacyjny 4	Zestaw walidacyjny 5	Zestaw walidacyjny 6	Zestaw walidacyjny 7	Zestaw walidacyjny 8	Zestaw walidacyjny 9	Zestaw walidacyjny 10
Przesunięcie czasowe [ms]	65	30	-5	-25	-20	75	115	-40	55	80
						Średnie przesunięcie czasowe [ms]			33	

Jak widać, sieć oparta na klasyfikatorze impulsów osiągała bardzo dobre, w zasadzie 100% poprawne wyniki klasyfikacji oraz nieznacznie lepszą terminowość wykrycia w porównaniu z analizowanym wcześniej podejściem opartym o paczki danych.

W celu potwierdzanie możliwości pracy modelu jako klasyfikatora online dokonano integracji sieci ze zbudowaną wcześniej aplikacją LabVIEW i sterownikiem PXIe. Środowisko programistyczne pozwoliło na bezpośrednie zastosowanie narzędzia wykrywania jako części kodu źródłowego systemu akwizycji danych. Aplikacja opierała się o układ pomiarowy 2 (z kartą PXIe-6363) i używała sygnału z generatora EDM do deterministycznego określania granic kolejnych impulsów. W czasie pracy algorytm prowadził akwizycję danych, na ich podstawie klasyfikował każdy z impulsów, przekazywał dane o ich liczebności do modelu CNN-GRU, który wykrywał przebicie. Dane pomiarowe, etykiety impulsów oraz flaga wykrywania przebicia zapisywane były w plikach wyjściowych. W celu potwierdzenia prawidłowości działania algorytmu, mierzono czasy wykonania głównych wewnętrznych pętli programowych. Dodatkowo dokonano kilkugodzinnych testów syntetycznych, gdzie na wejście karty pomiarowej zadawano znane przebiegi napięciowe z generatora funkcyjnego. Jako wynik testu porównywano terminowość danych pomiarowych oraz inferencji etykiety impulsu i przebicia w różnych obszarach plików danych wyjściowych. Testy potwierdziły, że dzięki znacznemu zmniejszeniu złożoności klasyfikatora impulsów jest on w stanie prowadzić inferencję podczas prowadzenia procesu, bez opóźnień. Co za tym idzie całość modelu będzie w stanie prowadzić terminowe wykrywanie przebicia tym podejściem.

5.2.3. Testy wykrywania przebicia

Proponowane metody wykrywania przebicia otworu zostały zastosowane w walidacyjnych eksperymentach wiercenia otworów przelotowych. Zbudowane narzędzia zostały zintegrowane z systemem pomiarowym 2 (jak opisano w rozdziale 5.2.2). Aby zweryfikować metodę, wywiercono po 20 otworów z wykorzystaniem każdego z modeli dla różnych zestawów parametrów. Do systemu pomiarowego dodano funkcjonalność integrującą maszynę MKG 1000 z systemem pomiarowo-kontrolnym. Opracowane algorytmy i nauczone wcześniej modele zostały wykorzystane bezpośrednio w kodach źródłowych aplikacji pomiarowych tworzonych w środowisku LabVIEW. System, w momencie wykrycia przebicia generował sygnał cyfrowy w standardzie TTL z karty ePIX-6363, który trafiał na wejście sterownika elektrodrążarki. W sterowniku maszyny zaprogramowany odpowiednie G-kody (odpowiedzialne za zatrzymanie ruchu maszyny) oraz M-kody (odpowiedzialne za zatrzymanie generacji iskier) pozwalające na szybką reakcję na wykrycie przebicia.

Otrzymanie sygnału powodowało zatrzymanie procesu wiercenia (zatrzymanie posuwu i generacji iskier). Testy prowadzono na zestawie 2 płytek Inconel718 oddalonych od siebie o odległość 1mm. Poprawność wykrycia zakończenia każdego z otworów określano poprzez oględziny obu płytek zestawu. Przy poprawnym działaniu algorytmu płytka 1 powinna posiadać w pełni przelotowy otwór o jednolitej średnicy, a na płytce 2 nie powinno być obserwowalnej erozji materiału oraz dane pomiarowe nie powinny wskazywać na ponowne rozpoczęcie generacji iskrzenia (brak zajścia błędu „overdrillingu"). Zestawienie wyników pokazano w poniższej tabeli.

Tabela 24 Wyniki testów wykrywania przebicia opracowanymi metodami

Model wykrywania przebicia	Zestaw parametrów	Ilość otworów [szt.]	Orientacja elektrody	Wykryte przebicia	Przedwczesne wykrycia	Oznaki erozji płytki 2
CNN (oparty o zestawy danych)	Parametry #1 (optymalne)	20	prostopadła	20	0	0
CNN (oparty o zestawy danych)	Parametry #1 (optymalne)	20	kątowa	20	0	0
CNN (oparty o zestawy danych)	Parametry #2 (agresywne)	20	prostopadła	20	0	0
CNN (oparty o zestawy danych)	Parametry #2 (agresywne)	20	kątowa	20	0	0
CNN-GRU (oparty o klasyfikację impulsów)	Parametry #1 (optymalne)	20	prostopadła	20	0	0
CNN-GRU (oparty o klasyfikację impulsów)	Parametry #1 (optymalne)	20	kątowa	20	0	0
CNN-GRU (oparty o klasyfikację impulsów)	Parametry #2 (agresywne)	20	prostopadła	20	0	0
CNN-GRU (oparty o klasyfikację impulsów)	Parametry #2 (agresywne)	20	kątowa	20	0	0

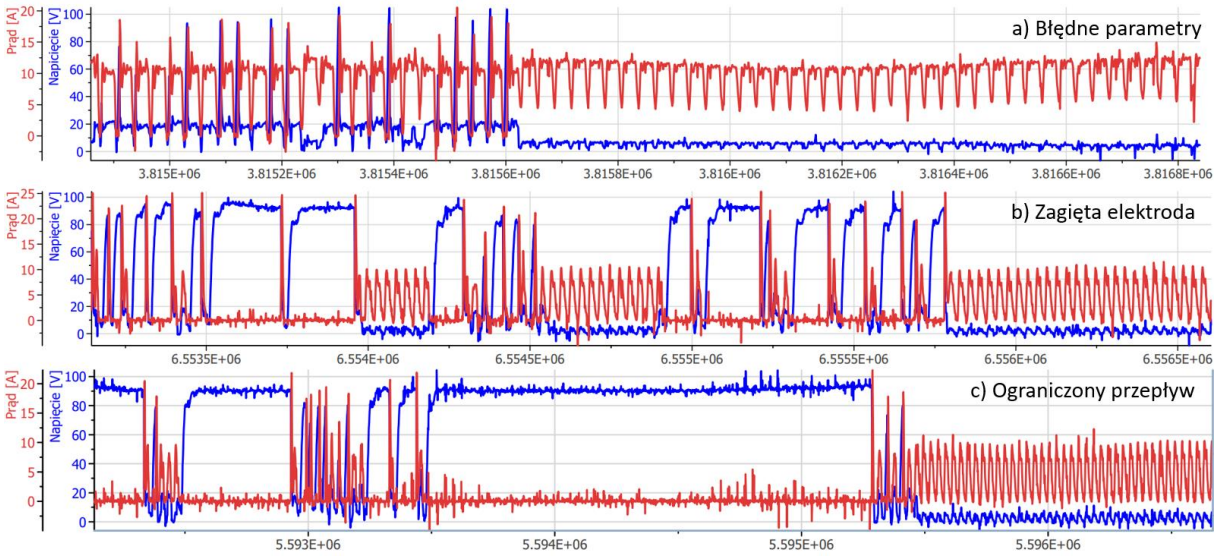
Wyniki eksperymentalne pokazują, że proponowana metoda może bardzo skutecznie wykrywać przebicia. W drążeniach eksperymentalnych, poprawność wykrywania osiągnęła 100%. W żadnym z wierceń nie zaobserwowano erozji na dolnych płytkach zestawu testowego. Wyniki potwierdziły skuteczność i praktyczność proponowanej metody.

5.3. Wykrywanie anomalii procesu FHD

Jak opisywano w rozdziale 3.5 metody wykrywania anomalii mogą być pomocnym narzędziem przy analizie wyników drążeń procesu tworzenia otworów chłodzenia filmowego. Podczas analizy literaturowej dokonano przeglądu metod stosowanych do automatycznego identyfikowania anomalii. Niestety metody te nie były do tej pory stosowane podczas analizy procesów EDM. W niniejszym rozdziale dokonam opisu metodologii, która pozwoliła na opracowanie algorytmu pozwalającego na skuteczne identyfikowanie anomalii FHD.

Jako część wierceń testowych wykonano szereg drążeń, w których zasymulowano niezgodności obróbki FHD: błędne parametry obróbki, wiercenie wygiętą elektrodą, ograniczenie przepływu dielektryka i zjawisko scarfingu. Symulowane błędy są przykładami usterek jakie zidentyfikowano na wcześniejszym etapie prac badawczo rozwojowych technologii wytwarzania otworów chłodzących. Dane z tych wierceń oraz wiercenia z dwoma

zestawami parametrów (optymalnymi #1 i agresywnymi #2) posłużyły jako dane wejściowe rozpoznawania anomalii. Celem pracy było opracowanie metody, która z dużą dokładnością będzie w stanie rozpoznawać dane anomalne i jednoznacznie odróżniać je od danych powstających podczas normalnego wiercenia z parametrami optymalnymi #1.

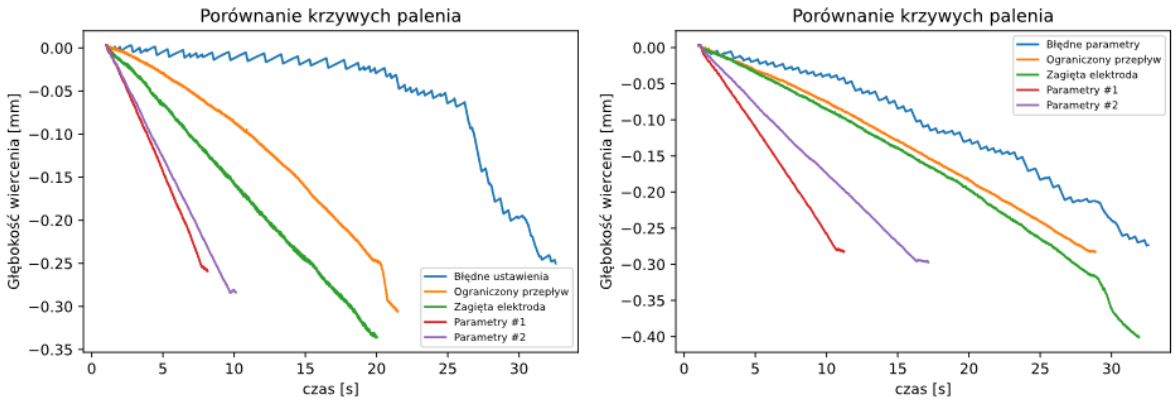


Rysunek 93 Przykładowe przebiegi prądu i napięcia wierceń anomalnych. a) parametry błędne, b) zagięta elektroda, c) graniczony przepływ

Przegląd danych przebiegów symulowanych awarii ujawnił, że zadanie rozpoznawania anomalii nie jest zadaniem trywialnym. Na rysunku 93 przedstawiono przykładowe fragmenty przebiegów prądu i napięcia dla danych anomalnych.

Jedynie wiercenia z parametrami błędnymi wykazywały wyraźne różnice w porównaniu z danymi poprawnymi. W przebiegach tych zidentyfikować można było w zasadzie występowanie jedynie różnej formy wylądowań łukowych oraz zwarć występujących naprzemiennie z długimi okresami obwodów otwartych zachodzących najprawdopodobniej w wyniku wycofywania wrzeciona po wykryciu stanu zwarciowego. Obserwowane wyładowania łukowe dodatkowo charakteryzowały się innymi przebiegami (w tym większymi amplitudami prądu) niż wyładowania łukowe występujące podczas normalnego drążenia. Niestety przebiegi prądu i napięcia dla zagiętych elektrod, ograniczonego przepływu i scarfingu posiadały kształt zbliżony do normalnych. Główną zidentyfikowaną różnicą była częstotliwość występowania poszczególnych impulsów, w tym zwiększona relatywna ilością zwarć dla wygiętych elektrod i ograniczonego przepływu. Takie przebiegi utrudniały jednoznaczne identyfikowanie anomalii.

Ze względu na różne warunki prowadzenia procesu można było rozróżniać wiercenia analizując je w skali makro. Jedną z metod analizy jakości wierceń otworów jest przegląd tzw. krzywych palenia. Metoda ta polega na wykreślaniu relatywnego przesunięcia wrzeciona roboczego wzdłuż osi Z w funkcji czasu wiercenia. Przykładowe krzywe palenia dla wierceń z dwoma zestawami parametrów i stanami anomalnymi przedstawiono na rysunku 94.



Rysunek 94 Porównanie krzywych palenia dla różnych wierceń testowych

Tak powstała krzywa daje informację o tempie obróbki i częściowo o stabilności procesu. Poprawne wiercenia powinny charakteryzować się krótkim czasem obróbki oraz liniowym przebiegiem zagłębiania elektrody. Odstępstwa od linii prostej sugerują zmienność warunków drążenia lub częste wycofywanie elektrody likwidujące zwarcia. Przedstawione przebiegi pokazują fragment drążenia od momentu rozpoczęcia drążenia do momentu wykrycia przebicia płytki testowej.

Wiercenia z dwoma zestawami parametrów #1 i #2, mają liniowe przebiegi i znacznie krótszy czas obróbki od wierceń anomalnych. Oba zestawy parametrów mają liniowe przebiegi krzywych, co świadczy o relatywnie dużej stabilności wiercenia. Stany anomalne wykazują dużo większe czasy obróbki oraz mniej stabilne warunki pracy. Dla zagiętej elektrody zachodzą miejscowe odstępstwa od idealnie liniowego przebiegu, dla ograniczonego przepływu dodatkowo zaobserwować można wyraźną nieliniowość krzywej. Wiercenie z parametrami błędnymi bardzo wyraźnie odbiegają od normy. Widać dużo następujących po sobie odcinków szybkiego zagłębiania elektrody z widocznym wycofywaniem elektrody. Zwrócić uwagę należy także na fakt wykrycia zakończenia wiercenia na różnych głębokościach dla różnych zestawów parametrów. Każde z wierceń wykonywane było na detalach o tej samej grubości, co pozwala domniemywać, że różnica w zagłębieniu elektrody dla różnych drążeń wynika z różnego tempa zużywania elektrody. Widać, że parametry optymalne nie tylko kończą

drążenie najszybciej, ale także zużywają najmniej elektrody roboczej. Parametry agresywne #2 są mniej optymalne pod tymi względami. Zgodnie z przebiegiem krzywych, zagięcie elektrody powoduje też jej intensywne zużywanie. Odpowiada to obserwowanym wynikam drążeń. Elektrody zagięte w efekcie wykonują drążenie otworów o zmienionej geometrii tzn. powiększonej średnicy niż zewnętrzna średnica elektrody. Fakt ten jest bezpośrednią przyczyną większego erodowania elektrody.

Analiza krzywych palenia w pewnym stopniu może być narzędziem pozwalającym na wykrywanie stanów anomalnych. Niestety analiza taka może być prowadzona tylko w skali makro. Dla przykładu, analiza taka może nie wykrywać krótkotrwałych błędów wiercenia, które ze względu na czas trwania nie będą widoczne na krzywych, ale dalej mogą mieć negatywny wpływ np. na miejscową jakość powierzchni wykonanego otworu.

Kolejną z metod analizy przynoszącą pewne informacje o przebiegu wiercenia, tym razem już w skali mikro, może być analiza rozkładu impulsów prowadzona na podstawie zbudowanego w ramach rozprawy klasyfikatora. Podejście to może pozwalać na identyfikowanie obserwowanych w przebiegach czasowych zmienionych częstotliwości zachodzenia wyładowań różnego typu. Podejście to niestety ma poważne ograniczenia. W wypadku wyraźnej zmiany charakteru samych impulsów jak w wypadku całkowitej zmiany zestawu parametrów maszyny, klasyfikator impulsów może nie działać poprawnie. W efekcie podejście takie może prowadzić do błędnych wniosków. Z tego względu, konieczne jest opracowanie niezależnej metody rozpoznającej anomalie. W tym celu przetestowano kilka podejść z tej dziedziny opisywanych podczas przeglądu literatury w rozdziale 3.5.

Biorąc pod uwagę doświadczenia wyciągnięte z wcześniejszych etapów prac prowadzonych w ramach rozprawy, rozpoznawanie anomalii procesu oparto o analizę pakietów danych składających się z 10 000 kolejnych pomiarów wartości prądu i napięcia. Podejście takie pozwala na terminowe wykrywanie anomalii (krótkie czasy analizy – pakiety danych po 20ms) oraz pozwala niwelować krótkookresowo zmienny charakter przebiegów.

Podejścia klasyczne lasów izolacyjnych i OC-SVM uczą się rozpoznawać różnice pomiędzy danymi poprawnymi i anomaliami, w efekcie zwracając binarną etykietę stanu. Ponieważ obie metody zakładają izolowanie znanej ilości anomalii już na etapie uczenia, przygotowano zestawy uczące składające się z dużej ilości danych poprawnych z wiercenia z parametrami #1, dodatkowo zanieczyszczonych znaną ilością anomalii (około 10% całej

populacji) z każdej z symulowanych grup oraz danymi z drążeń wykonanych z parametrami #2. Na takich zestawach danych dokonano testów. Istotnym warunkiem poprawności wykrywania anomalii, oprócz skuteczności w rozpoznawaniu danych odstających, jest także minimalizacja błędnych identyfikacji (błąd II rodzaju). Dlatego optymalizacja parametrów wykrywania anomalii oboma metodami była prowadzona do momentu osiągnięcia przynajmniej 98% poprawnej klasyfikacji stanów normalnych.

Stosując metodę lasów izolacyjnych zoptymalizowano współczynnik zanieczyszczenia (contamination) do poziomu zapewniającego 98% rozpoznawalność danych poprawnych. Wyniki rozpoznawania anomalii tą metodą z danych uczących i dwóch niezależnych zestawów walidacyjnych przedstawia Tabela 25.

Tabela 25 Wyniki wykrywania anomalii metodą lasów izolacyjnych

Isolation Forest contamination =0,228	Dane poprawne (parametry #1)	Parametry Błędne	Elektroda zagięta	Ograniczony przepływ	Scarfing	Dane parametry #2
Dane uczące	98%	95%	95%	98%	95%	17%
Dane walidacyjne 1	99%	91%	98%	100%	84%	23%
Dane walidacyjne 2	98%	93%	87%	99%	93%	34%

Lasy losowe dość dobrze radziły sobie z wykrywaniem różnych klas anomalii. Dane walidacyjne także dawały powtarzalne wyniki klasyfikacji. Algorytm z dobrą celnością oznaczał dane z wierceń z parametrami błędnymi, ograniczonym przepływem, zagiętą elektrodą i scarfingiem jako anomalie procesu. Niestety, metoda ta nie dawała sobie rady z rozpoznawaniem przebiegów stosujących zestaw parametrów agresywnych #2. W tej grupie zaledwie 17% całej populacji zestawów danych została poprawnie oznaczona jako anomalie. Wynika to z dużego podobieństwa przebiegów drążeń z parametrami #1 i #2, różnią się one nieznacznie długością okresów, obciążeniem cyklu i częstością impulsów.

Kolejną testowaną metodą było OC-SVM z jądrem liniowym. Zestawienie wyników pokazuje poniższa tabela.

Tabela 26 Wyniki wykrywania anomalii metodą OC-SVM (jądro liniowe)

SVM linear kernel nu=0,2	Dane poprawne (parametry #1)	Parametry Błędne	Elektroda zagięta	Ograniczony przepływ	Scarfing	Dane parametry #2
Dane uczące	98%	72%	50%	70%	26%	0%
Dane walidacyjne 1	99%	49%	99%	100%	24%	23%
Dane walidacyjne 2	99%	47%	94%	99%	21%	34%

Liniowe OC-SVM wyraźnie gorzej radziło sobie z rozpoznawaniem anomalii. Metoda ta w danych uczących nie przekraczała 70% poprawności klasyfikacji. Podobnie do lasów izolacyjnych najmniejszą skuteczność osiągała dla rozpoznawania parametrów agresywnych, niską skuteczność zanotowano także dla skarfingu. Warto zwrócić także uwagę na dużą zmienność wyników klasyfikacji dla danych walidacyjnych.

Sprawdzono także wariant OC-SVM z jądrem nieliniowym w postaci radialnej funkcji bazowej RBF.

Tabela 27 Wyniki wykrywania anomalii metodą OC-SVM (jądro rbf)

SVM rbf kernel nu=0,18	Dane poprawne (parametry #1)	Parametry Błędne	Elektroda zagięta	Ograniczony przepływ	Scarfing	Dane parametry #2
Dane uczące	98%	99%	99%	98%	76%	45%
Dane walidacyjne 1	98%	64%	49%	63%	69%	57%
Dane walidacyjne 2	99%	64%	63%	62%	77%	68%

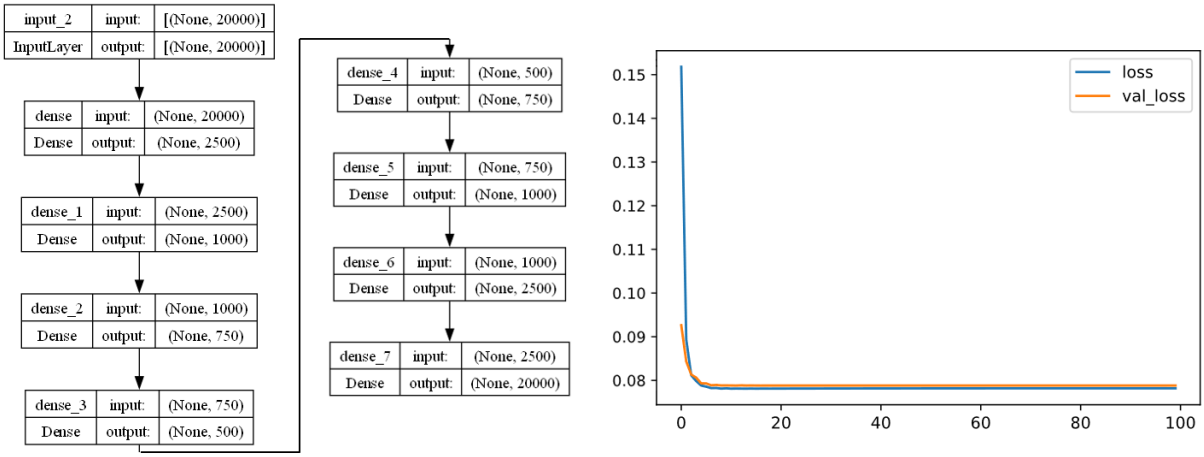
Dla danych uczących, wariant rbf znacznie lepiej poradził sobie z każdym typem anomalii. Osiągnął też najlepszy spośród testowanych metod współczynnik rozpoznawalności parametrów #2. Niestety testy na danych walidacyjnych pokazały małą pewność algorytmu. Metoda dawała znacznie gorsze wyniki niż na danych uczących.

Spośród przetestowanych klasycznych metod wykrywania anomalii lasy izolacyjne wykazywały najlepszą skuteczność w wykrywaniu anomalii oraz odznaczały się powtarzalnymi wynikami dla różnych danych walidacyjnych. Niestety, jak opisywano powyżej metoda ta nie radzi sobie z odróżnianiem drążeń wykonywanych z różnymi zestawami parametrów. Testy wykazały, że zmiany parametrów każdej z metod może zwiększać wykrywalność anomalii, ale zawsze kosztem większej ilości błędnych klasyfikacji danych poprawnych jako anomalie.

Wyniki osiągnięte klasycznymi metodami wykrywania anomalii nie spełniły wymogów postawionych jako cel w rozprawie, dlatego rozpoznano kolejne podejście: podejście rekonstrukcyjne. W podejściu tym opracowana metoda najpierw uczy się jak najlepszej reprezentacji danych poprawnych i wykrywa anomalie na podstawie współczynnika błędu rekonstrukcji między nową daną wejściową, a sygnałem zrekonstruowanym.

W pierwszym podejściu opracowano wielowarstwowy autoenkoder oparty o gęste sieci neuronowe MLP. Autoenkoder posiadał strukturę symetrycznego niedopełnionego enkodera i dekodera z warstwą wąskiego gardła jako warstwy je łączącej. Enkoder dokonywał

zmniejszenia wymiarowości danych wejściowych i ekstrakcji cech, zadaniem dekodera było jak najlepsze rekonstruowanie sygnału do wymiarowości wejściowej. Podobnie jak w wypadku modeli opisywanych w poprzednich rozdziałach, sprawdzono różne konfiguracje autoenkodera stosując metody przeszukiwanie siatki i przeszukiwania losowego. Strukturę ostatecznie zbudowanej sieci oraz krzywą straty uczenia i walidacji pokazuje poniższy rysunek 95.

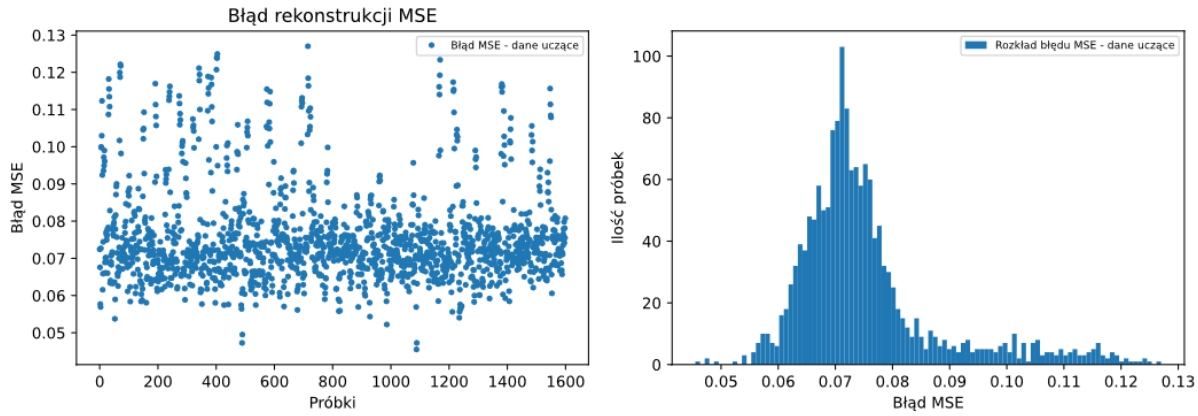


Rysunek 95 Struktura autoenkodera MLP wykrywania anomalii i funkcja straty uczenia i walidacji

Optymalne wyniki osiągnięto dla autoenkodera wielowarstwowego z 3 warstwami ukrytymi o zmniejszającej się ilości neuronów z funkcjami aktywacji ReLU. Wąskie gardło stanowiła warstwa gęsto połączona o 500 neuronach z sigmoidalną funkcją aktywacji. Dekoder zbudowany był jako symetryczne odbicie enkodera. Warstwę wyjściową stanowiła warstwa gęsto połączona z sigmoidalną funkcją aktywacji. Najlepszy proces uczenia dawało używanie danych normalizowanych, zastosowanie optymalizatora Adam i funkcji błędu średniokwadratowego MSE (ang. Mean Square Error). Ostateczna wartość funkcji straty wynosiła 0,0755 dla danych uczących i 0,0752 dla danych walidacyjnych. Ze względu na dużą ilość danych wejściowych oraz kilka gęstych warstw ukrytych, wyjściowy model posiadał bardzo złożoną budowę. W modelu było aż 119 033 000 uczonych parametrów i zajmował on 454,07 MB pamięci.

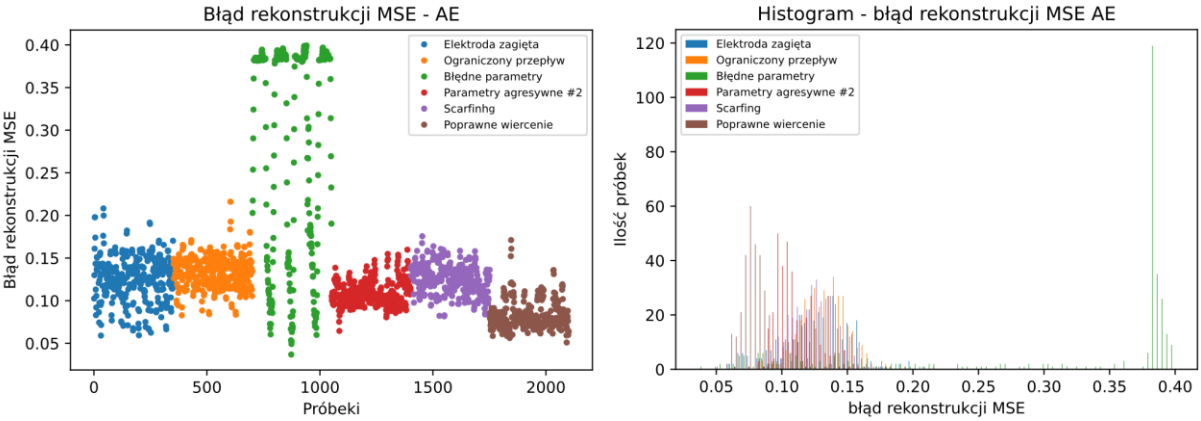
Na tak zbudowanym autoenkoderze dokonano rekonstrukcji zestawu danych testowych osiągając maksymalny błąd rekonstrukcji na poziomie 0,13. Jako miarę błędu rekonstrukcji wybrano błąd średniokwadratowy obliczany pomiędzy całymi zestawami danych wejściowych, a odpowiadającymi im zestawami danych zrekonstruowanych. Wybór MSE podyktowany był czułością tej metody na wartości odstające. W porównaniu do średniego absolutnego błędu,

MSE powinno mocniej eksponować anomalie. Na rysunku 96 przedstawiono wykres wartości błędu rekonstrukcji dla każdego zestawu testowego oraz histogram rozkładu błędu.



Rysunek 96 Błąd rekonstrukcji i histogram błędu dla danych uczących AE MLP

W kolejnym kroku na wejście autoenkodera podano specjalnie przygotowany zestaw danych testowych złożony z 350 pakietów dla każdej grupy danych anomalnych różnych kategorii i danych poprawnych, kolejno: elektroda zagięta, ograniczony przepływ, błędne parametry, parametry #2, scarfing i parametry poprawne. Rysunek 97 przedstawia wartości błędu rekonstrukcji dla każdego pakietu danych walidacyjnych oraz histogram rozkładu tych błędów.

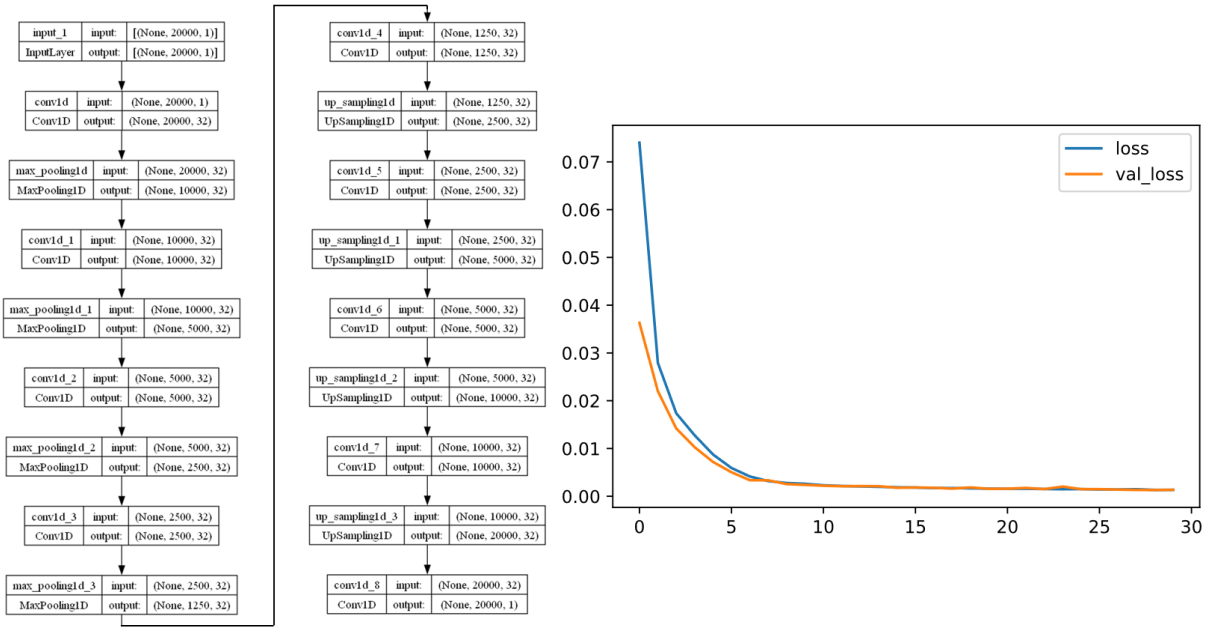


Rysunek 97 Wartości i histogram średniokwadratowych błędów rekonstrukcji dla danych walidacyjnych (danych anomalnych)

Analizując wyniki rekonstrukcji, widzimy pewne, obserwowalne różnice w wartościach błędów MSE pomiędzy różnymi grupami przebiegów anomalnych, a wiercenie zasadniczym. Niestety istnieje bardzo duża grupa pakietów danych anomalnych, które odznaczają się błędem rekonstrukcji w zakresie normalnych błędów rekonstrukcji danych poprawnych. Widać to dobrze na histogramie błędu, rozkłady błędów anomalnych bardzo mocno przenikają

obszar rozkładu błędu wierceń poprawnych. Z tego względu nie da się ustalić jednej, granicznej wartości błędu MSE rekonstrukcji, który pozwalałby na jednoznaczne identyfikowanie anomalii.

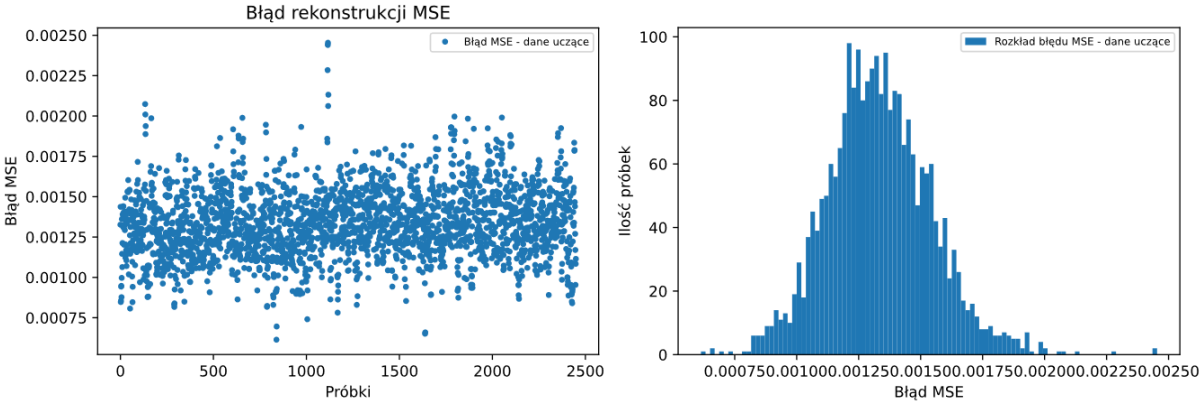
W poprzednich badaniach, udowodniono, że podejście sieci konwolucyjnych CNN 1D ma możliwość rozpoznawania przebiegów prądu i napięcia EDM. Zdolność CNN 1D do ekstrakcji różnych zestawów cech, może dawać możliwość rozpoznawania, nie tylko różnic w kształcie samych impulsów, ale także różnic w częstotliwości ich zachodzenia w całych pakietach danych napięcia i prądu. Daje to podstawę do inwestygacji podejścia opartego o sieci konwolucyjne, także do wykrywania anomalii. W dalszych badaniach zaproponowano zbudowanie wielowarstwowego, niedopełnionego autoenkodera CNN 1D jako narzędzia rozpoznawania błędów. W tym celu sprawdzono różne konfiguracje budowy takich enkoderów. Ostateczną architekturę sieci oraz krzywe funkcji straty uczenia i walidacji przedstawiono poniżej.



Rysunek 98 Struktura autoenkodera CNN 1D wykrywania anomalii i funkcja straty uczenia i walidacji

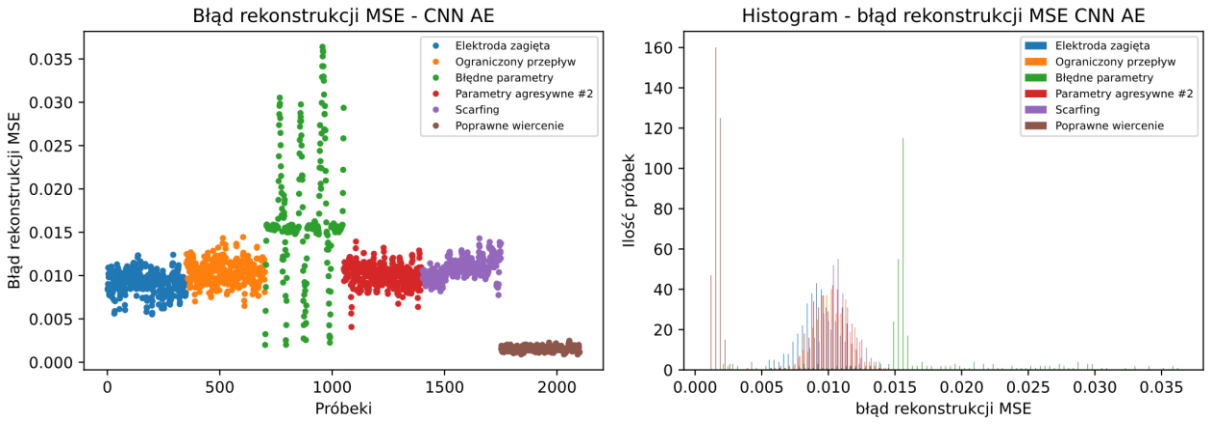
Optymalne wyniki osiągnięto dla enkodera wielowarstwowego z 4 konwolucyjnymi warstwami ukrytymi z aktywacją ReLU z następującymi za nimi warstwami max poolingu. W warstwach konwolucyjnych zastosowano jądro o długości 32 próbek oraz 32 filtry. Wąskie gardło stanowiła warstwa konwolucyjna przetwarzająca sygnały o długości 1250 próbek. Dekoder stanowił symetryczne odzwierciedlenie enkodera, z warstwami up samplingu zamiast warstw max poolingu. Warstwę wyjściową stanowiła warstwa konwolucuyjna z sigmoidalną

funkcją aktywacji i jednym filtrem. Najlepszy proces uczenia dawało używanie danych normalizowanych, zastosowanie optymalizatora Adam i funkcji błędu średniokwadratowego MSE (ang. Mean Square Error). Ostateczna wartość funkcji straty wynosiła 0,013 dla danych uczących i 0,0014 dla danych walidacyjnych, co jest wynikiem znacznie lepszym niżwynik osiągnięty w podejści wielowarstwowego AE MLP. Dodatkową zaletą odejścia opartego o sieci CNN jest mniejsza złożoność modelu. Zbudowany autoenkoder posiadał tylko 231 681 uczonych parametrów i zajmował 905 KB pamięci (ponad 500 razy mniej niż w podejściu AE MLP).



Rysunek 99 Błąd rekonstrukcji i histogram błędu dla danych uczących AE CNN 1D

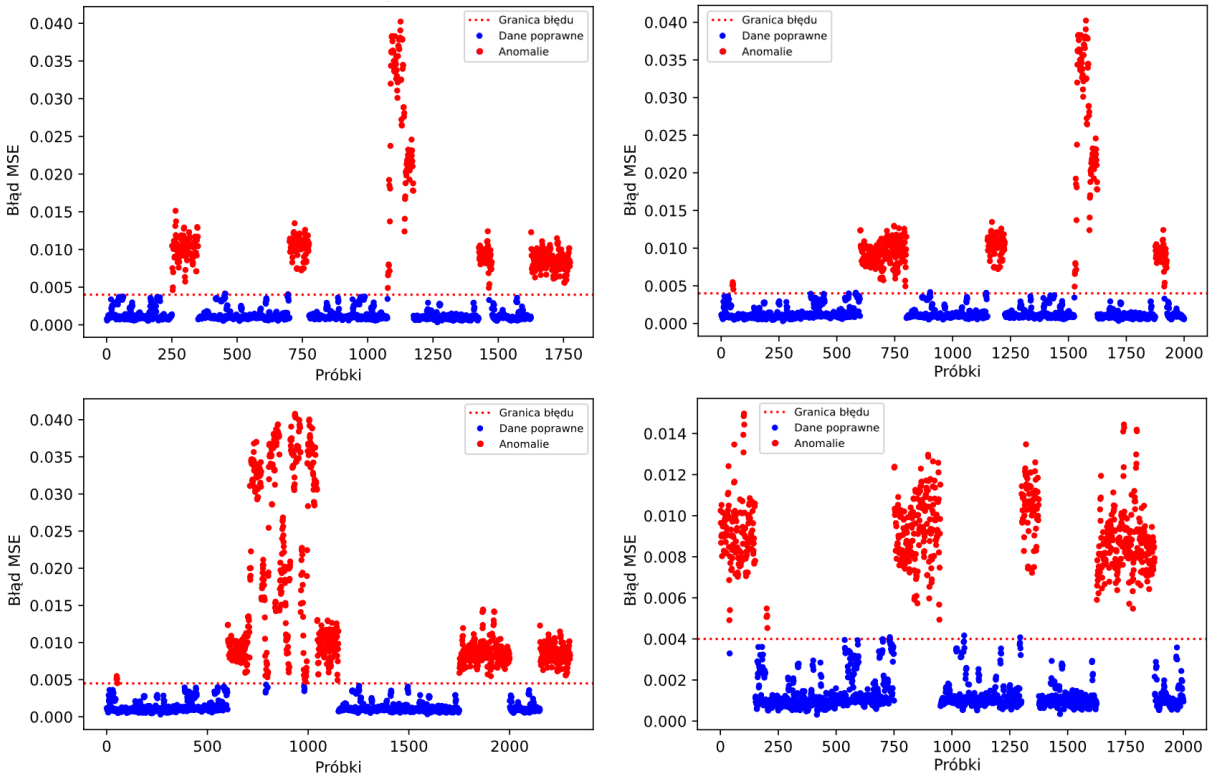
Jak pokazuje powyższy rysunek 99 już na etapie uczenia, autoenkoder oparty o sieci konwolucyjne dawał znacznie lepsze wyniki osiągając niższe wartości średniokwadratowego błędu rekonstrukcji dla danych uczących. Pozwalało to domniemywać, że tak zbudowany autoenkoder może stanowić dobre narzędzie do wyławiania anomalii z danych napięcia i prądu.



Rysunek 100 Wartości i histogram średniokwadratowych błędów rekonstrukcji dla danych walidacyjnych (danych anomalnych)

Podobnie jak w wypadku testów Autoenkodera opartego o MLP, przeanalizowano błędy rekonstrukcji dla danych walidacyjnych złożonych z pakietów danych anomalnych. Zestawienie wyników pokazano na rysunku 100. Zastosowanie podejścia konwolucyjnego dramatycznie poprawiło zdolności rekonstrukcyjne autoenkodera. Enkoder posiada też wyraźną zdolność do rozróżniania przebiegów anomalnych, a nawet przebiegów wykonanych z innym zestawem parametrów. Analizując histogram rozkładu błędów MSE widzimy mocne skupienie danych normalnych w okolicy małych wartości błędów oraz rozkłady anomalne posiadające wyraźnie większe błędy. Poszczególne grupy parametrów są prawie całkowicie rozdzielne. Oznacza to, że można określić jedną wartość progową dla średniokwadratowego błędu rekonstrukcji, po przekroczeniu której można jednoznacznie stwierdzić, że dany pakiet danych jest sytuacją anomalną. Warto zwrócić uwagę na błędy rekonstrukcji osiągnięte dla parametrów błędnych. Jak pokazują wykresy, w tej grupie istnieje pewna niewielka cześć danych, które odznaczają się błędem rekonstrukcji zbliżonym do danych normalnych. Po dokładniejszym przeglądzie, ustalono, że próbki te stanowią zestawy danych złożone prawie z samych obwodów otwartych z czasów wycofywania elektrody po wykryciu zwarcia. W drążeniach wykonywanych z użyciem parametrów #1, okresowo występują długie obszary obwodów otwartych. Ze względu na bardzo duże podobieństwo tych przebiegów dane takie mogą być klasyfikowane przez autoenkoder blisko granicy anomalii. Sytuacja taka nie jest krytyczna, ponieważ wystąpienie dużej ilości obwodów otwartych samo w sobie nie powinno powodować uszkodzeń detalu obrabianego.

Jak pokazały wyniki podejście rekonstrukcyjne oparte o sieci CNN jest skutecznym narzędziem do wykrywania anomalii procesu FHD. Aby ostatecznie potwierdzić tą tezę, przeprowadzono analizy na dużej ilości przebiegów syntetycznych stworzonych z naprzemiennie występujących odcinków złożonych z danych poprawnych, z wplecionymi w nie odcinakami danych anomalnych złożonych ze wszystkich kategorii wierceń błędnych oraz danych z drążeń wykonanych z zestawem parametrów #2. Dla analizy określono poziom granicznej wartości błędu rekonstrukcji. Każdy pakiet danych wykazujący błąd rekonstrukcji poniżej wartości granicznej klasyfikowany był jako dana poprawna, każdy powyżej granicy oznaczany był jako anomalia. Na rysunku 101 przedstawiono przykładowe przebiegi testowe.



Rysunek 101 Wyniki identyfikacji anomalii na podstawie AE CNN 1D

Przegląd wyników tak przeprowadzonych testów ostatecznie potwierdził skuteczność metody. Aby określić ostateczną skuteczność metody stworzono duży plik danych walidacyjnych złożony z danych poprawnych nie biorących udziału w procesu uczenia autoenkodera oraz zestawu danych błędnych. Każda z kategorii sygnałów miała taką samą liczebność z zestawie testowym. Dane błędne pobrane były z losowych czasów drążenia z różnych drążeń testowych symulujących błędy procesu. Otrzymane wyniki poprawności rozpoznawalności anomalii przedstawiono w poniższej tabeli.

Tabela 28 Wyniki rozpoznawalności anomalii z użyciem rekonstrukcyjnego CNN AE

	Parametry poprawne	Parametry błędne	Zagięta elektroda	Ograniczony przepływ	Scarfing	Parametry nr#2
Wykrywalność anomalii [%]	100,00%	96,85%	100,00%	100,00%	100,00%	100,00%
				Średnia skuteczność [%]		99,48%

Przy odpowiednim dobraniu granicy błędu rekonstrukcji algorytm rozpoznawania anomalii osiągał bardzo dobre wyniki. Ponad 99,48% wszystkich danych walidacyjnych została rozpoznana prawidłowo. Jedynie dane z wierceń z parametrami błędnymi, ze względu na przyczyny opisywane wcześniej posiadały niedużą ilość przebiegów zaklasyfikowanych

jako poprawne. Nadal, prawie 97% danych z tej kategorii zostało oznaczonych prawidłowo. Wszystkie dane z wierceń wykonanych z dobrymi parametrami zostało zakwalifikowane do danych poprawnych. Tak samo 100% sygnałów danych symulujących błędy wynikających z zagiętej elektrody, scarfingu, ograniczonego przepływu i wierceń z parametrami #2 zostało poprawnie oznaczone jako anomalie. Osiągnięte wyniki należy uznać za bardzo dobre. Przeprowadzone badania potwierdziły możliwość zastosowania podejścia rekonstrukcyjnego opartego o autoenkoder CNN 1D jako skutecznego narzędzia rozpoznawania sytuacji błędnych i potencjalnie niebezpiecznych dla procesu wiercenia FHD otworów wentylacyjnych.

5.4. Wnioski

Opisane w rozprawie badania i testy pozwoliły na zbudowanie kilku skutecznych narzędzie diagnozowania i kontroli procesu FHD. W ramach prac zbudowano skuteczne narzędzia automatycznej klasyfikacji impulsów EDM osiągające skuteczność przekraczającą 99,6% oraz dokonano analizy statystycznej i przeglądu zmienności występowania impulsów w zależności od etapu procesu. Na bazie tych doświadczeń zaproponowano, zbudowano i przetestowano dwie metody wykrywania przebicia (skuteczność 100%) oraz klasyfikator anomalii (wykrywalność anomalii 99,48%). Prace prowadzone były z zastosowanie metodologii ucznia maszynowego i uczenia głębokiego, szczególnie z wykorzystaniem głębokich sieci neuronowych. Doświadczenia pracy potwierdziły potencjał tych narzędzi, także w procesie FHD. Możliwość pracy na surowych danych, zdolność do uogólniania wnioskowania i zmniejszenie nakładu pracy inżynierskiej w znacznym stopniu ułatwiły i przyśpieszyły procesy badawcze. W pracy udowodniono także, że metody sztucznej inteligencji stanowią nie tylko narzędzie do analizy danych offline, ale z zastosowaniem odpowiednich podejść i optymalizacji, mogą także być silnymi rozwiązaniami do kontroli procesów w czasie rzeczywistym. Podejście takie dało dobre wyniki nawet w przetwarzaniu wysokoczęstotliwościowych sygnałów prądu i napięcia jakie zachodzą przy drążeniu EDM.

6. Podsumowanie przeprowadzonych prac badawczych

W niniejszej rozprawie opisane zostało proponowane podejście do analizy i kontroli procesu FHD. Proces ten jest obecnie główną metodą wykonywania otworów wentylacyjnych filmowego chłodzenia wysokotemperaturowych elementów silników lotniczych. Stale prowadzone są prace mające na celu rozwój tej technologii. Z tego względu opracowane rozwiązania mają duży potencjał wdrożeniowy w przemysłowym produkowaniu tych części.

Podczas całego procesu badań i rozwoju przeprowadzonego w trakcie przygotowywania rozprawy osiągnięto następujące cele:

- Identyfikacja problemów procesu FHD oraz szczegółowy przegląd literatury tematu
- Stworzenie systemu pomiarowego zbierania wysokoczęstotliwościowych sygnałów elektrycznych procesu
- Wykonanie serii wierceń testowych z różnymi parametrami oraz serii wierceń symulujących błędy procesu FHD
- Szczegółowy przegląd sygnałów procesowych w skali makro i mikro. Identyfikacja charakteru typowych impulsów wiercenia
- Opracowanie rozwiązania do automatycznej synchronizacji danych z różnych systemów (poprawna kompensacja przesunięć czasowych rzędu 70, 5894 μs/s)
- Stworzenie narzędzia automatycznego podziału impulsów na kategorie w zależności od ich charakteru.
- Opracowanie niezawodnego klasyfikatora impulsów opartego o sieci neuronowe. Testy rozwiązania oraz porównanie różnych metod ML i DL. Wyłonienie optymalnego rozwiązania osiągającego dokładność klasyfikacji na poziomi 99,6%
- Charakteryzacja częstości wystąpień oraz analiza statystyczna różnych klas impulsów w zależności od parametrów oraz etapu procesu

- Testy różnych metod i podejść do wykrywania przebicia otworu. Wyłonienie dwóch optymalnych podejść, opartych o uczenie głębokie, osiągających 100% poprawności i terminowości wykrywania.

- Testy różnych metodologii wykrywania anomalii w analizie przebiegów prądu i napięcia FHD. Opracowanie i testy skutecznego narzędzia opartego o głęboki konwolucyjny autoenkoder, osiągającego skuteczność rozpoznawania anomalii na poziomie 99,48%

W pracy opisano podstawowe założenia procesu oraz zidentyfikowano główne trudności z jakimi boryka się obecnie przemysł, głównie zagadnienie identyfikacji przebicia oraz wykrywanie błędów i anomalii procesu. W części analizy literaturowej opisano obecny stan wiedzy w temacie przeciwdziałania tym problemom oraz dokonano przeglądu publikacji opisujących charakter wysokoczęstotliwościowych sygnałów elektrycznych, odpowiedzialnych za generację wyładowań EDM. Obecnie proponowane rozwiązania nie zapewniają w pełni niezawodnego rozwiązania wykrywania przebicia. Nie zidentyfikowano także żadnych prac podejmujących próbę rozpoznawania sytuacji anomalnych i błędów drążenia bezpośrednio na podstawie przebiegów prądu i napięcia EDM. Opublikowane do tej pory prace analizujące proces FHD w niewielkim stopniu korzystały z potencjału metodologii opartych o ML i DL, a żadna z nich nie podjęła próby wdrożenia tych metod do kontroli przebiegu procesu w czasie rzeczywistym. Dlatego tezą pracy było wdrożenie szeroko pojętych metodologii opartych o sztuczną inteligencję i uczenie maszynowe, jako narzędzi dających szerokie możliwości w analizie charakteru procesu oraz jego diagnostyki. W dalszej części pracy dokonano szerokiego przeglądu stosowalności metod ML i DL w zagadnieniach diagnostyki procesów przemysłowych i wykrywania anomalii. Analiza aktualnego stanu wiedzy, pozwoliła na utwierdzenie przekonania o potencjale użycia narzędzi opartych ideę sztucznej inteligencji do kontroli i wykrywania niezgodności w procesie wiercenia otworów FHD.

W części badawczej opisano rozwiązanie systemu pomiarowego do akwizycji wysokoczęstotliwościowych sygnałów prądu i napięcia EDM. Z pomocą systemu wykonano serię drążeń eksperymentalnych, które posłużyły jako podstawa dalszej analizy danych.

Na podstawie zebranych danych dokonano kompleksowego badania charakteru przebiegów czasowych sygnałów w skali makro (całych drążeń) oraz w skali mikro identyfikując przebiegi pojedynczych wyładowań EDM procesu FHD.

W dalszej części zaprezentowano skuteczne rozwiązania ML i DL do podziału impulsów na kategorie w zależności od przebiegu wyładowań. Opracowany model osiągał dokładność przekraczającą 99,6% poprawnie sklasyfikowanych impulsów. Rozwiązania te w pełni korzystały z potencjału zastosowanych metod, dzięki czemu osiągnięto dużą automatyzację pracy i zdolność wnioskowania modeli bezpośrednio z nieobrobionych danych pomiarowych. Pozwoliło to na szczegółowy opis procesu i da możliwość lepszego zrozumienia wpływu stosowanych parametrów obróbki na rzeczywisty przebieg procesu.

Zaprojektowano także nowatorskie podejście do wykrywania zakończenia formowania otworu podczas wiercenia FHD. W pracy opisano testy metod ML i DL oraz zaproponowano dwa niezależna podejścia do tego zagadnienia. Proponowane metody działają w czasie rzeczywistym podczas procesu i korzystają z surowych danych pomiarowych, bez konieczność dodatkowej ekstrakcji cech z sygnałów oraz minimalizują potrzebny nakład pracy inżynierskiej wymaganej w procesie przygotowawczym. Przeprowadzone badania udowodniły wysoką, praktycznie stuprocentową poprawność oraz terminowość wykrywalności etapu procesu. Obie z proponowanych metod mają potencjał wdrożeniowy w procesie produkcyjnym.

Ostatnią częścią pracy jest opracowanie metodologii automatycznego wykrywania anomalii procesu. Założeniem pracy było opracowanie narzędzia pozwalającego na niezawodne identyfikowanie usterek procesu oraz automatyczne rozpoznawanie różnic wynikających ze stosowania innych parametrów obróbki na postawie przebiegów prądu i napięcia. Testy opracowanego rozwiązania pozwoliły na rozpoznawanie ponad 99,48% spośród różnych błędnych danych symulowanych podczas wierceń testowych. Rozwiązanie pozwala także na zachowanie wysokiej rozdzielczości analizy dając możliwość poprawnej interpretacji anomalii w czasie tak krótkim jak 20ms.

Podsumowując, w ramach rozporowy opracowano oraz przetestowano szereg narzędzi i metod pozwalających na usprawnienie prowadzenia procesu FHD oraz dających potencjał na ich zastosowanie w przyszłych pracach badawczo rozwojowych technologii. Prace w dużej mierze opierały się o metodologie sztucznej inteligencji w szczególności głębokie sieci neuronowe. Pomimo małej gamy publikacji tematu stosujących takie podejście, doświadczenia

eksperymentalne potwierdziły stawianą na początku pracy tezę o dużym potencjale takich podejść, także w diagnostyce i kontroli procesu FHD. Proponowana metodologia korzysta z surowych danych pomiarowych procesu, nie wymaga pracochłonnego ekstrahowania cech i minimalizuje nakład pracy ekspertów. Zaproponowane narzędzia zostaną następnie wdrożone w procesie produkcyjnym i zostaną przetestowane na faktycznych częściach silników odrzutowych.

7. Dalsze prace badawcze

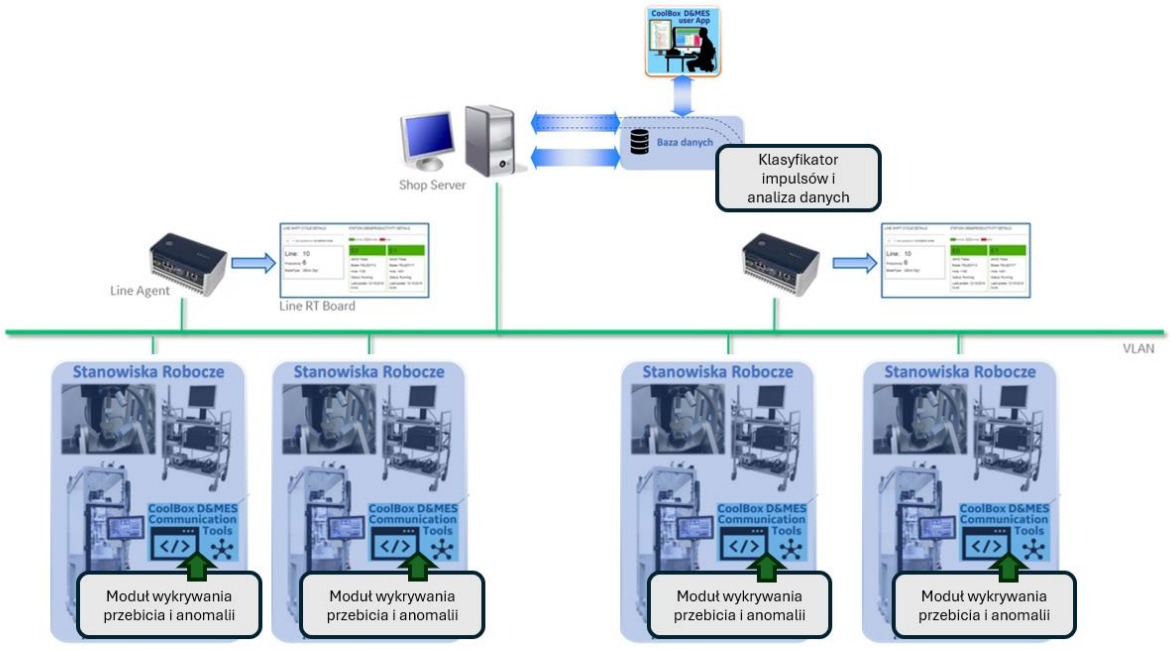
W rozprawie opisywano badania przeprowadzone z użyciem materiałów i metod reprezentujących prawdziwe części lotnicze. Kolejnym krokiem w pracach badawczych będzie przetestowanie zaproponowanych metodologii podczas prowadzenia prawdziwego procesu produkcyjnego. Konieczne będzie przeprowadzenie szeroko zakrojonych testów drążeń części turbin lotniczych. Ideą testów będzie zebranie danych z realnego procesu i dostrojenie modeli wykrywania przebicia i wykrywania anomalii z wykorzystaniem metodologii uczenia transferowego (ang. transfer learning). Podejście takie powinno pozwolić na szybkie wdrożenie technologii. Podczas testów zaplanowane jest także dokonanie optymalizacji modeli. Jak pokazały testy, wykrycie przebicia i zatrzymanie procesu cechuje się obecnie pewnymi rozbieżnościami w czasie zadziałania w zależności od przebiegu samego procesu. Spowodowanie to było strojeniem modeli na podstawie przewidywanych czasów przebicia. Wykonanie długiej serii testów, w których zatrzymanie drążenia będzie wykonywane przez opracowany algorytm, pozwoli na korelację długości elektrody wystającej za ściankę detalu po zatrzymaniu z momentem przebicia przewidzianym przez sieć neuronową. Wykonanie dużej ilości takich korelacji pozwoli na poprawę terminowości i powtarzalności działania metody.

Kolejnym planowanym zadaniem będzie integracji opracowanych metod z obrabiarką. Obecnie trwają prace nad przeniesieniem opracowanych narzędzi na platformy sprzętowe pozwalające na łatwą integrację z elektrodrążarką MKG. Opracowany system będzie pozwalał na zbieranie wysokoczęstotliwościowych danych, przetwarzanie i wnioskowanie na ich podstawie w czasie trwania procesu. Dodatkowym ograniczeniem jest także konieczność opracowania rozwiązania optymalnego od strony kosztowej, ponieważ system w przyszłości może być powielony w środowisku produkcyjnym złożonym z wielu pracujących równolegle maszyn. Częścią prac integracyjnych jest także opracowanie aplikacji interfejsu użytkownika, który będzie pozwalał operatorowi maszyny na łatwą interakcję z systemem wykrywania przebicia i anomalii. Trwają także prace nad aplikacją ekspercką wizualizującą i ułatwiającą etapy przygotowania danych i uczenia sieci. Założeniem jest stworzenie HMI (ang. Human Machine Interface), które będzie pozwalało na wizualizację danych uczących i prowadzenie procesu optymalizacji sieci w przystępnej formie graficznej.

Kolejnym kierunkiem działań jest integracja z systemami bazodanowymi wspierającymi procesy produkcyjne. Dane zbierane przez system oraz wyniki analiz będą archiwizowane w

bazie danych systemu MES (ang. Manufacturing Execution System). Zastosowanie już zaimplementowanych w systemie metod analizy danych, pozwoli na szerszą analizę nowych danych o przebiegach sygnałów elektrycznych, charakterze impulsów EDM, ich przebiegach oraz informacji o zidentyfikowanych anomaliach procesu. Podejście takie może otworzyć nowe możliwości korelacji wyników drążeń z czynnikami takimi jak parametry procesu, stosowane materiały czy dokładności geometryczne narzędzi. Analiza anomalii pozwoli także na łatwiejsze identyfikowanie przyczyn powstawania defektów obrabianych detali i może posłużyć jako narzędzie optymalizacji procesów kontroli jakości.

W przyszłości istnieje możliwość rozszerzenia stosowalności opracowywanych systemów poza środowisko laboratoryjne i wdrożenie ich w produkcji seryjnej. W tym wypadku system zostanie przystosowany do pracy w warunkach przemysłowych i może ideowo wyglądać jak na poniższym rysunku 102.



Rysunek 102 Przykładowa architektura systemu w środowisku produkcyjnym

W warunkach produkcyjnych konieczne będzie zintegrowanie rozwiązań akwizycji danych i inferencji przebicia oraz anomalii na każdej z maszyn oraz zbudowanie infrastruktury sieciowej pozwalającej na komunikację każdej z maszyn z centralną bazą danych systemu CoolBox.

Jednym z kierunków rozwojowych technologii drążenia otworów wentylacyjnych jest zastąpienie dielektryka, którym płukana jest szczelina przez specjalnie przygotowany

roztwór soli. Zmiana taka w zasadzie powoduje przejście z czystego procesu elektroerozyjnego do procesu stanowiącego połączenie obróbki EDM i obróbki elektrochemicznej ECM. Obrabiarka EDM MKG 1000 została już wzbogacana o system przygotowania roztworu soli o określonych parametrach. Pierwsze kampanie drążeń testowych przyniosły obiecujące rezultaty. Zmiana warunków prowadzenie procesu będzie wymagała przeprowadzenia szeroko zakrojonych badań, mających na celu optymalizację parametrów i lepsze zrozumienie różnic w porównaniu z czystym EDM. Już na starcie tego procesu zakładane jest korzystanie z opracowanych w ramach tej rozprawy narzędzi. Możliwość zbierania wysokoczęstotliwościowych danych prądu i napięcia oraz ich przetwarzania pod względem przebiegów wyładowań, będą stanowiły wartościowe informacje i mogą znacznie przyśpieszyć i wzbogacić procedury badawcze. Ze względu na istotną zmianę fizycznego charakteru powstawania procesu erozji elektrody i materiału obrabianego (dodanie czynnika elektrochemicznego), procedury opracowania klasyfikatorów przebicia oraz anomalii powinny zostać przeprowadzone ponownie, w celu ich dostosowania do zmienionych przebiegów procesu.

8. Bibliografia

1. Tortorella G., Miorando R., Caiado R., D. Nascimento, Portioli Staudacher A., (2021). "The mediating effect of employees' involvement on the relationship between Industry 4.0 and operational performance improvement", Total Quality Management & Business Excellence, 32(1–2): 119–133. DOI: 10.1080/14783363.2018.1532789

2. PWC (2016), "Global industry 4.0 survey: building the digital enterprise", https://www.pwc.com/gx/en/industries/industries-4.0/landing-page/industry-4.0-building-your-digital-enterprise-april-2016.pdf (dostęp 14.01.2024).

3. Hermann M., Pentek T., Otto B. (2016), "Design principles for industrie 4.0 scenarios", IEEE 49th Hawaii International Conference on System Sciences, 3928-3937. DOI10.1109/HICSS.2016.488

4. Yang C., Shen W., Wang X. (2016), "Applications of Internet of Things in manufacturing", IEEE 20th International Conference on Computer Supported Cooperative Work in Design (CSCWD), 670–675, DOI:10.1109/CSCWD.2016.7566069

5. Zhong R.Y., Ge W., (2018), "Internet of things enabled manufacturing: a review", International Journal of Agile Systems and Management 11(2):126–154 DOI:10.1504/IJASM.2018.10013695

6. Yang C., Shen W., Wang X., (2018), "The internet of things in manufacturing: key issues and potential applications", IEEE Systems, Man, and Cybernetics Magazine 4(1): 6–15, DOI: 10.1109/MSMC.2017.2702391

7. Lee J., Davari H., Singh J., Pandhare V. (2018) "Industrial artificial intelligence for Industry 4.0-based manufacturing systems", Manufacturing Letters 18:20–23. DOI: 10.1016/j.mfglet.2018.09.002

8. Li B.H., Hou B.C., Yu W.T., Lu X.B., Yang C.W. (2017), "Applications of artificial intelligence in intelligent manufacturing: a review". Frontiers of Information Technology & Electronic Engineering 18(1):86–96, DOI:10.1631/FITEE.1601885

9. Menezes, B. C., Kelly, J. D., Leal, A. G. (2019), “Identification and Design of Industry 4.0 Opportunities in Manufacturing: Examples from Mature Industries to Laboratory Level Systems”. IFAC-PapersOnLine, 52(13): 2494-2500, DOI: 10.1016/j.ifacol.2019.11.581

10. Siderska J., Jadaan K.S. (2018), “Cloud manufacturing: a service oriented manufacturing paradigm. A review paper”. Engineering Management in Production and Services, 10(1):22–31, DOI:10.1515/emj-2018-0002

11. Yan J., Meng Y., Lu L., Guo C. (2017), “Big-data-driven based intelligent prognostics scheme in industry 4.0 environment”, 2017 Prognostics and System Health Management Conference (PHM-Harbin): 1-5, DOI:10.1109/PHM.2017.8079310 IEEE.

12. Huang D. C., Lin C. F., Chen C. Y., Sze, J. R. (2018), “The Internet technology for defect detection system with deep learning method in smart factory”, 2018 4th International Conference on Information Management (ICIM): 98-102, DOI: 10.1109/INFOMAN.2018.8392817

13. Wang J., Ma Y., Zhang L., Gao R. X., Wu D. (2018), “Deep learning for smart manufacturing: Methods and applications”, Journal of Manufacturing Systems, 48: 144156, DOI: 10.1016/J.JMSY.2018.01.003

14. Womack J.P., Jones D.T. (19976), “Lean thinking - banish waste and create wealth in your corporation”, Journal of the Operational Research Society, 48(11), DOI: 10.1038/sj.jors.2600967

15. Liker J.K., Franz, J.K. (2011), “The Toyota Way to Continuous Improvement: Linking Strategy and Operational Excellence to Achieve Superior Performance”, McGraw-Hill, New York, NY, ISBN: 978-0-07-147746-8

16. Kanigolla D., Cudney E.A., Corns S.M., Samaranayake V.A. (2014), “Enhancing engineering education using project-based learning for Lean and Six Sigma”, International Journal of Lean Six Sigma, 5(1): 45-61, DOI:10.1108/IJLSS-02-2013-0008

17. Buer S.V., Strandhagen J.O., Chan, F.T. (2018), “The link between Industry 4.0 and lean manufacturing: mapping current research and establishing a research agenda”, International Journal of Production Research, 56(8):2924-2940, DOI: 10.1080/00207543.2018.1442945

18. Rossini M., Costa F., Tortorella G.L., Portioli-Staudacher A. (2019), “The interrelation between Industry 4.0 and lean production: an empirical study on European manufacturers”, The International Journal of Advanced Manufacturing Technology, 102(9-12):3963-3976, DOI: 10.1007/s00170-019-03441-7

19. Tao F., Qi Q., Liu A., Kusiak A. (2018), “Data-driven smart manufacturing”. Journal of Manufacturing Systems, 48: 157–169, DOI: 10.1016/j.jmsy.2018.01.006

20. Nasir V., Sassani F. (2021), “A review on deep learning in machining and tool monitoring: methods, opportunities, and challenges”, International Journal of Advanced Manufacturing Technology, 115(9-10): 2683-2709, DOI: 10.1007/s00170-021-07325-7

21. Teti R., Jemielniak K., O’Donnell G., Dornfeld D. (2010), „Advanced monitoring of machining operations”. CIRP Annals – Manufacturing Technology, 59(2):717–739, DOI: 10.1016/j.cirp.2010.05.010

22. Abellan-Nebot J.V., Subirón F.R. (2010), “A review of machining monitoring systems based on artificial intelligence process models”. International Journal of Advanced Manufacturing Technology, 47(1-4):237–257, DOI: 10.1007/s00170-009-2191-8

23. Lauro C.H., Brandão L.C., Baldo D., Reis R.A., Davim J.P. (2014), “Monitoring and processing signal applied in machining processes–a review”. Measurement, 58:73–86, DOI: 10.1016/j.measurement.2014.08.035

24. Zhu K., San Wong Y., Hong G.S. (2009), “Wavelet analysis of sensor signals for tool condition monitoring: a review and some new results”. International Journal of Machine Tools & Manufacture, 49(7-8):537–553, DOI: 10.1016/j.ijmachtools.2009.02.003

25. Kusiak A. (2019), “Fundamentals of smart manufacturing: a multithread perspective”, Annual Reviews in Control 47:214–220, DOI: 10.1016/j.arcontrol.2019.02.001

26. The Technical Publications Department, (1988), “The Jet Engine”. Derby, UK: Rolls-Royce plc,. ISBN 0902121 235.

27. Bunker R. S. (2017), “Evolution of turbine cooling”, ASME Turbo Expo 2017: Turbomachinery Technical Conference and Exposition, DOI: 10.1115/GT2017-63205

28. Sims C. T. (1984), “A history of superalloy metalurgy for superalloy metalurgists”, Superalloys, 399-419, DOI: 10.7449/1984/SUPERALLOYS_1984_399_419.

29. Cumpsty N. (2003), “Jet Propulsion: A Simple Guide to the Aerodynamics and Thermodynamics Design and Performance of Jet Engines”, Cambridge University Press, ISBN-10: 0521541441

30. Bunker R. S., Dees J. E, Palafox P, (2014), “Impingement Cooling in Gas Turbines: Design, Applications, and Limitations”, WIT Transactions on State of the Art in Science and Engineering, 76:1755–8336, doi:10.2495/978-1-84564-907-4/001

31. Bunker R.S. (2005), “A review of shaped hole turbine film cooling technology”, Journal of Heat Transfer, 127:441-453, DOI: 10.1115/1.1860562

32. Pusavec F., Deshpande A., Yang S., M’Saoubi R., Kopac J., Dillon O., Jawahir I (2014), “Sustainable Machining of High Temperature Nickel Alloy—Inconel 718: Part 1-Predictive Performance”, Models Journal of Cleaner Production, 81: 255–269, DOI: 10.1016/j.jclepro.2014.06.040

33. Devillez A., Le Coz G., Dominiak S., Dudzinski D. (2011), “Dry Machining of Inconel 718, Workpiece Surface Integrity”, Journal of Materials Processing Technology 211(10): 1590–1598, DOI: 10.1016/j.jmatprotec.2011.04.011

34. Dudzinski D., Devillez A., Moufki A., Larrouquere V., Vigneau J. (2004), “A review of developments towards dry and high speed machining of Inconel 718 alloy”, International Journal of Machine Tools and Manufacture 44(4): 439–456, DOI: 10.1016/S0890-6955(03)00159-7

35. Hao Z., Gao D., Fan Y., Han R. (2011), “New observations on tool wear mechanism in dry machining Inconel 718”, International Journal of Machine Tools and Manufacture 51(12): 973–979, DOI: 10.1016/j.ijmachtools.2011.08.018

36. Ulutan D., Ozel T. (2011), “Machining induced surface integrity in titanium and nickel alloys: a review”, International Journal of Machine Tools and Manufacture 51(3): 250-280, DOI: 10.1016/j.ijmachtools.2010.11.003

37. Attia H., Tavakoli S., Vargas R., Thomson V. (2010), “Laser-assisted high-speed finish turning of superalloy Inconel 718 under dry conditions”, CIRP Annals – Manufacturing Technology, 59(1): 83–88, DOI10.1016/j.cirp.2010.03.093

38. Pušavec F., Kramar D., Krajnik P., Kopac J. (2010), “Transitioning to sustainable production—part II: evaluation of sustainable machining technologies”, Journal of Cleaner Production 18(12): 1211–1221, DOI10.1016/j.jclepro.2010.01.015

39. Ezugwu E.O. (2005), “Key improvements in the machining of difficult-to-cut aerospace superalloys”, International Journal of Machine Tools and Manufacture 45(12-13): 1353-1367, DOI: 10.1016/j.ijmachtools.2005.02.003

40. Wang Z.Y., Rajurkar K.P., Fan J., Lei S., Shin Y.C., Petrescu G. (2003), “Hybrid machining of Inconel 718”, International Journal of Machine Tools and Manufacture 43(13): 1391–1396, DOI: 10.1016/S0890-6955(03)00134-2

41. Ezugwu E.O., Bonney J., Fadare D.A., Sales W.F. (2005), “Machining of nickel-base, Inconel 718, alloy with ceramic tools under finishing conditions with various coolant supply pressures”, Journal of Materials Processing Technology, 162: 609–614, DOI: 10.1016/j.jmatprotec.2005.02.144

42. Elman J. C. (2001), “Electrical Discharge Machining”, Dearborn: Society of Manufacturing Engineers, ISBN: 0-87263-521-X

43. Schumaker M. (2013), “Historical phases of EDM development driven by the dual Influence of “Market Pull” and “Science Push””, Seventeenth CIRP Conference On Electro Physical And Chemical Machining (ISEM), 6(5-12), DOI: 10.1016/j.procir.2013.03.001.

44. Liang W., Tong H., Li Y., Li B., Kong Q. (2019), “Sliding-mode controller and algorithm for improving servo control of discharge gap in precise fast hole EDM”, International journal of advanced manufacturing technology, 105(5-6): 2689-2698, DOI: 10.1007/s00170-019-04481-9

45. Kumar K. (2017), “Optimizing the Electrical Discharge Drilling Process for High Aspect Micro Hole Drilling in Die Steel”, Handbook of Research on Manufacturing Process Modeling and Optimization Strategies, 120-141, DOI: 10.4018/978-1-5225-2440-3.ch006

46. Colban W., Thole K. (2007), “Influence of hole shape on the performance of a turbine vane endwall film-cooling scheme”, International Journal of Heat and Fluid Flow, 28, (3): 341-356, DOI: 10.1016/j.ijheatfluidflow.2006.05.002

47. Lee KD, Kil SM, Kim KY (2013), “Multi-Objective Optimization of a Row of Film Cooling Holes Using an Evolutionary Algorithm and Surrogate Modeling”, Numerical heat transfer part A-Applications, 63(8): 623-641, DOI: 10.1080/10407782.2013.751316

48. Gostimirovic M., Kovac P., Sekulic M., Skoric B. (2012), “Influence of discharge energy on machining characteristics in EDM”, Journal of Mechanical Science and Technology 26(1): 173-179, DOI: 10.1007/s12206-011-0922-x

49. Liao YS, Woo JC. (1997), “The effects of machining settings on the behavior of pulse trains in the WEDM process”, Journal of Materials Processing Technology, 71(3): 433-439, DOI: 10.1016/S0924-0136(97)82076-6

50. Weck M., Dehmer J. M. (1992), “Analysis and Adaptive Control of EDM Sinking Process Using the Ignition Delay Time and Fall Time as Parameter”, CIRP Annals, 41(1): 243-246, DOI: 10.1016/S0007-8506(07)61195-0

51. Caggiano A., Teti R., Perez R., Xirouchakis P. (2015), “Wire EDM Monitoring for Zero-Defect Manufacturing based on Advanced Sensor Signal Processing”, 9th CIRP Conference on Intelligent Computation in Manufacturing Engineering - CIRP ICME’14, 33: 315-320, DOI: 10.1016/j.procir.2015.06.065

52. Oßwald K., Lochmahr I., Schulze H. P., Kröningb O. (2018), “Automated Analysis of Pulse Types in High Speed Wire EDM”, 19th CIRP Conference on Electro Physical and Chemical Machining, 68: 796 – 801, DOI: 10.1016/j.procir.2017.12.157

53. Zhou M., Wu J., Yang J., Yao D. (2016), “Fast and Stable Electrical Discharge Machining (EDM) by Two-step-ahead Predicted Control”, 18th CIRP Conference on Electro Physical and Chemical Machining (ISEM XVIII), 42: 215 – 220, DOI: 10.1016/j.procir.2016.02.274

54. Zhou M., Wu, J. Xu X., Mu X., Dou Y. (2018), “Significant improvements of electrical discharge machining performance by step-by-step updated adaptive control laws”,

Mechanical Systems and Signal Processing, 101: 480–497, DOI: 10.1016/j.ymssp.2017.06.041

55. Hegab H. A., Gadallah M. H., Esawi A. K. (2015), “Modeling and optimization of Electrical Discharge Machining (EDM) using statistical design”, Manufacturing Review, 2(21), DOI: 10.1051/mfreview/2015023

56. Gangil M., Pradhan M. K. (2017), “Modeling and optimization of electrical discharge machining process using RSM: A review”, Materials Today-Proceedings, 4(2): 1752-1761,

57. Laxman J., Raj K. G. (2015), “Mathematical modeling and analysis of EDM process parameters based on Taguchi design of experiments”, Vibration Problems, 662, DOI: 10.1088/1742-6596/662/1/012025

58. Sánchez H. T., Estrems M., Faura F. (2011), “Development of an inversion model for establishing EDM input parameters to satisfy material removal rate, electrode wear ratio and surface roughness”, International Journal of Advanced Manufacturing Technology, 57(1-4):189–201, DOI: 10.1007/s00170-011-3283-9

59. Sciaroni B. (1988), “Process and apparatus for determining the electroerosive completion of a starting hole”, US Patent 4:767,903

60. Haefner K. B., Bischoff J. R., Ehresman M. D., Comeau S. F. (1994), “Controlled apparatus for electrical discharge machining”, US Patent 5:360,957

61. Yamada S., Takawashi T., Sakakibara T. (1984), “Breakthrough detection means for electric discharge machining apparatus”, U.S. Patent 4,484,051.

62. Koshy P., Boroumand M., Ziada Y. (2010), “Breakout detection in fast hole electrical discharge machining”, International Journal of Machine Tools & Manufacture, 50(10):922–925. DOI: 10.1016/j.ijmachtools.2010.05.006

63. Bellotti M., Qian J., Reynaerts D. (2020) “Self-tuning breakthrough detection for EDM drilling micro holes”. Journal Of Manufacturing Processes, 57:630–640, DOI: 10.1016/j.jmapro.2020.07.031

64. Lin J.K., Nien Y.F. (2004), “Automatic breakthrough detection device”, US Patent 6:723,942

65. Xia W., Li Z., Zhang Y., Zhao W. (2019), “Breakout detection for fast EDM drilling by classification of machining state graphs”. International Journal of Advanced Manufacturing Technology, 106(5-6):1645–1656, DOI: 10.1007/s00170-019-04530-3

66. Xia W., Wang J., Zhao W. (2018), “Break-out detection for high-speed small hole drilling EDM based on machine learning”. Procedia CIRP, 68:569–574. DOI: 10.1016/j.procir.2017.12.115

67. Zhang Y., Xia W., Li Z., Zhao W. (2021), “Completion detection and efficiency improvement for breakout stage of fast EDM drilling”, The International Journal of Advanced Manufacturing Technology, 114(5-6):1565-1574, DOI: 10.1007/s00170-021-06936-4

68. Bellotti M., Qian J., Reynaerts D. (2019), “Breakthrough phenomena in drilling micro holes by EDM”, International Journal of Machine Tools & Manufacture, 146, DOI: 10.1016/j.ijmachtools.2019.103436

69. Wang J., Xi X., Zhang Y., Zhao F., Zhao W. (2023), “Stage identification and process optimization for fast drilling EDM of film cooling holes using KBSI method”, Advenced Manufacturing, 11(3):477-491, DOI: 10.1007/s40436-022-00434-w

70. Liu R., Yang B., Zio E., Chen X. (2018), “Artificial intelligence for fault diagnosis of rotating machinery: A review”, Mechanical Systems and Signal Processing 108: 33–47, DOI: 10.1016/j.ymssp.2018.02.016

71. Duan L., Xie M., Wang J., Bai T. (2018), “Deep learning enabled intelligent fault diagnosis: Overview and applications”, Journal of Intelligent & Fuzzy Systems, 35(5): 5771– 5784, DOI: 10.3233/JIFS-17938

72. Cover T.M., Hart P.E. (1967), “Nearest neighbor pattern classification”, IEEE Transactions of Information Theory, 13(1): 21–27, DOI: 10.1109/TIT.1967.1053964

73. Schmidhuber J. (2015), “Deep learning in neural networks: An overview”, Neural Networks 61: 85–117, DOI: 10.1016/j.neunet.2014.09.003

74. Vapnik C. C. V. (1995), “Support-vector networks”, Machine Learning, 20: 273–297, DOI: 10.1023/A:1022627411411

75. Koller D., Friedman N. (2009), “Probabilistic Graphical Models: Principles and Techniques”, MIT Press, ISBN: 9780262013192

76. Knapp G.M., Wang H.P. (1992), “Machine fault classification - A neural network approach”, International Journal of Production Research, 30(4): 811–823, DOI: 10.1080/00207543.1992.9728458

77. Widodo A., Yang B.S. (2007), “Support vector machine in machine condition monitoring and fault diagnosis”, Mech. Syst. Signal Process. 21(6): 2560–2574, DOI: 10.1016/j.ymssp.2006.12.007

78. Ayvaz S., Alpay K. (2021), Predictive maintenance system for production lines in manufacturing: a machine learning approach using IoT data in real-time. Expert Systems with Applications, 173: 114598, DOI: 10.1016/j.eswa.2021.114598

79. Duan Z., Wu T., Guo S., Shao T., Malekian R., Li Z. (2018), “Development and trend of condition monitoring and fault diagnosis of multi-sensors information fusion for rolling bearings: A review”, Int. J. Adv. Manuf. Technol. 96(1-4): 803–819, DOI: 10.1007/s00170-017-1474-8

80. Lei Y., Zuo M.J., He Z., Zi Y. (2010), “A multidimensional hybrid intelligent method for gear fault diagnosis”, Expert Systems with Applications, 37(2): 1419–1430, DOI: 10.1016/j.eswa.2009.06.060

81. Lei Y., He Z., Zi Y., Hu Q. (2007), “Fault diagnosis of rotating machinery based on multiple ANFIS combination with gas”, Mechanical Systems and Signal Processing, 21(5): 2280–2294, DOI: 10.1016/j.ymssp.2006.11.003

82. Guyon I., Gunn S., Nikravesh M., Zadeh L. (2006), “Feature extraction, foundations and applications”, Springer, ISBN: 9783540354871

83. Metody redukcji wymiarowości przestrzeni cech — Sztuczna inteligencja, http://books.icse.us.edu.pl/runestone/static/ai/MetodyRedukcjiWymiarowosciPrzestrzeniCech/Wstep.html [dostęp 2024-01-20]

84. Duda R.O., Hart P.E., Stork D.G. (2000), Pattern Classification 2ed., Wiley, ISBN: 9780471056690

85. Chandrashekar G., Sahin F. (2014), "A survey on feature selection methods", Computers & Electrical Engineering 40(1):16–28, DOI: 10.1016/j.compeleceng.2013.11.024

86. Bolon-Canedo V., Sanchez-Marono N., Alonso-Betanzos A. (2013), "A review of feature selection methods on synthetic data", Knowledge and Information Systems, 34(3): 483-519, DOI: 10.1007/s10115-012-0487-8

87. Liu H., Setiono R. (1996), "Feature selection and classification - a probabilistic wrapper approach", International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems, 419–424

88. Hall M.A., Smith L.A. (1998), "Practical feature subset selection for machine learning", Proceedings of 21st Australasian Computer Science Conference ACSC'98, 20(1): 181-191

89. Kononenko I. (1994), "Estimating attributes: Analysis and extensions of RELIEF", European conference on machine learning, 171– 182, DOI: 10.1007/3-540-57868-4_57

90. Peng H., Long F., Ding C. (2005), "Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy", Ieee Transactions on Pattern Analysis and Machine Intelligence, 27(8): 1226–1238, DOI: 10.1109/TPAMI.2005.159

91. Yang B.S., Han T., An J. (2004), "ART-KOHONEN neural network for fault diagnosis of rotating machinery", Mechanical Systems and Signal Processing, 18(3): 645–657, DOI: 10.1016/S0888-3270(03)00073-6

92. Yang B.S., Kim K.J. (2006), "Application of Dempster-Shafer theory in fault diagnosis of induction motors using vibration and current signals", Mechanical Systems and Signal Processing. 20(2): 403–420, DOI: 10.1016/j.ymssp.2004.10.010

93. Tibshirani R., Saunders M., Rosset S., Zhu J., Knight K. (2005), "Sparsity and smoothness via the fused lasso", Journal of The Royal Statistical Society Series B-Statistical Methodology, 67: 91– 108, DOI: 10.1111/j.1467-9868.2005.00490.x

94. Tikhonov A.N., Goncharsky A.V., Stepanov V.V., Yagola A.G. (1995), “Numerical Methods for the Solution of Ill-Posed Problems”, Springer Ltd, DOI: 10.1007/978-94-015-8480-7

95. Musolf A. M., Holzinger E. R., Malley J. D., Bailey-Wilson J. E. (2022), “What makes a good prediction? Feature importance and beginning to open the black box of machine learning in genetics”, Human Genetics, 141(2): 1515-1528, DOI: 10.1007/s00439-021-02402-z

96. Glowacz A., Glowacz Z. (2017), “Diagnosis of stator faults of the single-phase induction motor using acoustic signals”, Applied Acoustics, 117: 20–27, DOI: 10.1016/j.apacoust.2016.10.012

97. Toma, R. N., Prosvirin A. E., Kim J. M. (2020), “Bearing Fault Diagnosis of Induction Motors Using a Genetic Algorithm and Machine Learning Classifiers”, Sensors, 20(7): 1884, DOI: 10.3390/s20071884

98. X. An, Y. Tang, Application of variational mode decomposition energy distribution to bearing fault diagnosis in a wind turbine, Trans. Inst. Meas. Control 39 (2017) 1000–1006.

99. Shinde P. V., Desavale R. G., Jadhav P. M., Sawant S. H. (2023), “A multi fault classification in a rotor-bearing system using machine learning approach”, Journal of The Brazilian Society of Mechanical Sciences And Engineering, 45(2): 121, DOI10.1007/s40430-023-04015-1

100. S. Park, S. Kim, J.H. Choi, Gear fault diagnosis using transmission error and ensemble empirical mode decomposition, Mech. Syst. Signal Process. 108 (2018) 262–275.

101. S.S. Vanraj, B.S. Dhami, Pabla, Hybrid data fusion approach for fault diagnosis of fixed-axis gearbox, Struct. Health Monit. 17 (2018) 936–945.

102. Kumar A., Parey A., Kankar P. K. (2023), “Supervised Machine Learning Based Approach for Early Fault Detection in Polymer Gears Using Vibration Signals”, Mapan-Journal of Metrology Society of India, 38(2): 383-394, DOI: 10.1007/s12647-022-00608-8

103. Zhao X., Jia M. (2018), “Fault diagnosis of rolling bearing based on feature reduction with global-local margin fisher analysis”, Neurocomputing 315: 447–464, DOI: 10.1016/j.neucom.2018.07.038

104. Dong S., Luo T., Zhong L., Chen L., Xu X. (2017), “Fault diagnosis of bearing based on the kernel principal component analysis and optimized k-nearest neighbour model”, Journal of Low Frequency Noise Vibration and Active Control, 36(4): 354–365, DOI: 10.1177/1461348417744302

105. Zhong Y., Wu X. (2020), “Effects of cost-benefit analysis under back propagation neural network on financial benefit evaluation of investment projects”, PloS ONE 15(3): e0229739, DOI: 10.1371/journal.pone.0229739

106. Rumelhart D.E., Hinton G.E., Williams R.J. (1986), “Learning representations by back-propagating errors”, Nature 323(6088): 533–536, DOI: 10.1038/323533a0

107. Merainani B., Rahmoune C., Benazzouz D., Ould-Bouamama B. (2018), “A novel gearbox fault feature extraction and classification using hilbert empirical wavelet transform, singular value decomposition, and som neural network”, J. Vibrat. Control 24(12): 2512–2531, DOI: 10.1177/1077546316688991

108. Chen X., Zhou J., Xiao H., Wang E., Xiao J., Zhang H. (2014), “Fault diagnosis based on comprehensive geometric characteristic and probability neural network”, Applied Mathematics and Computing, 230: 542–554, DOI: 10.1016/j.amc.2013.12.122

109. Shifat T. A., Hur J. (2021), “ANN Assisted Multi Sensor Information Fusion for BLDC Motor Fault Diagnosis”, IEEE ACCESS, 9: 9429-9441, DOI: 10.1109/ACCESS.2021.3050243

110. Barakat M., Lefebvre D., Khalil M., Druaux F., Mustapha O. (2013), “Parameter selection algorithm with self adaptive growing neural network classifier for diagnosis issues”, International Journal of Machine Learning and Cybernetics, 4(3): 217–233, DOI: 10.1007/s13042-012-0089-5

111. Chikkam S., Singh S. (2023) “Condition Monitoring and Fault Diagnosis of Induction Motor using DWT and ANN”, Arabian Journal for Science and Engineering, 48(5): 6237-6252, DOI10.1007/s13369-022-07294-3

112. Lopez C. (1999), “Looking inside the ANN “black box”: classifying individual neurons as outlier detectors”, IJCNN’99. International Joint Conference on Neural Networks. Proceedings (Cat. No. 99CH36339), 1185–1188, DOI: 10.1109/IJCNN.1999.831127

113. Wang Z., Fan J. (2015), “Fault early recognition and health monitoring on aeroengine rotor system”, Journal of Aerospace Engineering, 28(2): 04014065, DOI: 10.1061/(ASCE)AS.1943-5525.0000386

114. Pang B., Tang G., Zhou C., Tian T. (2018), “Rotor fault diagnosis based on characteristic frequency band energy entropy and support vector machine”, Entropy 20(12): 932, DOI: 10.3390/e20120932

115. Rapur J.S., Tiwari R. (2019), “On-line time domain vibration and current signals based multi-fault diagnosis of centrifugal pumps using support vector machines”, Journal of Nondestructive Evaluation, 38(1): 6, DOI: 10.1007/s10921-018-0544-7

116. Saravanan N., Siddabattuni V.N.S.K., Ramachandran K.I. (2008), “A comparative study on classification of features by SVM and PSVM extracted using morlet wavelet for fault diagnosis of spur bevel gear box”, Expert Systems with Applications, 35(3): 1351-1366, DOI: 10.1016/j.eswa.2007.08.026

117. Li Y., Feng K., Liang X., Zuo M.J. (2019), “A fault diagnosis method for planetary gearboxes under non-stationary working conditions using improved Vold-Kalman filter and multi-scale sample entropy”, Journal of Sound and Vibration, 439: 271–286, DOI10.1016/j.jsv.2018.09.054

118. Han H., Cui X. Y., Fan Y. Q., Qing H. (2019), “Least squares support vector machine (LS-SVM)-based chiller fault diagnosis using fault indicative features” Applied Thermal Engineering, 154: 540-547, DOI: 10.1016/j.applthermaleng.2019.03.111

119. Fong S., Narasimhan S. (2022), “An Unsupervised Bayesian OC-SVM Approach for Early Degradation Detection, Thresholding, and Fault Prediction in Machinery

Monitoring", IEEE Transactions on Instrumentation and Measurement, 71: 3500811, DOI: 10.1109/TIM.2021.3137858

120. Zhu Y., Du C., Liu Z., Chen Y., Zhao Y. (2022), "A Turboshaft Aeroengine Fault Detection Method Based on One-Class Support Vector Machine and Transfer Learning", Journal of Aerospace Engineering, 35(6): 04022085, DOI: 10.1061/(ASCE)AS.1943-5525.0001485

121. Dong S., Tang B., Chen R. (2013), "Bearing running state recognition based on non-extensive wavelet feature scale entropy and support vector machine", Measurement 46(10): 4189–4199, DOI: 10.1016/j.measurement.2013.07.011

122. Heidari M., Shateyi S. (2017), "Wavelet support vector machine and multi-layer perceptron neural network with continues wavelet transform for fault diagnosis of gearboxes", Journal of Vibroengineering, 19(1): 125–137, DOI: 10.21595/jve.2016.16813

123. Ghasemi S., Sadeghian A. (2021), "Application of Hybrid Wavelet-SVM Algorithm to Detect Broken Rotor Bars in Induction Motors", Proceedings Of 2021 Ieee 30th International Symposium on Industrial Electronics (ISIE), DOI10.1109/ISIE45552.2021.9576330

124. Zheng J., Pan H., Cheng J. (2017), Rolling bearing fault detection and diagnosis based on composite multiscale fuzzy entropy and ensemble support vector machines, Mech. Syst. Signal Process. 85 (2017) 746–759.

125. Li R., Ran C., Zhang B., Han L., Feng S. (2020), "Rolling Bearings Fault Diagnosis Based on Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise, Nonlinear Entropy, and Ensemble SVM", Applied Sciences-Basel, 10(16): 5542, DOI: 10.3390/app10165542

126. Hang J., Zhang J., Cheng M. (2016), "Application of multi-class fuzzy support vector machine classifier for fault diagnosis of wind turbine", Fuzzy Sets and Systems, 297: 128–140, DOI: 10.1016/j.fss.2015.07.005

127. Seegmiller C., Chamberlain B., Miller J., Masoum M. A. S., Shekaramiz M. (2022), "Wind Turbine Fault Classification Using Support Vector Machines with Fuzzy Logic",

2022 Intermountain Engineering, Technology and Computing (IETC), DOI10.1109/IETC54973.2022.9796919

128. Li Z., Zhong S., Lin L. (2017), “Novel gas turbine fault diagnosis method based on performance deviation model”, Journal of Propulsion and Power, 33(3): 730–739, DOI: 10.2514/1.B36267

129. Zhao X., Chen S., Gao K., Luo L. (2023), “Bidirectional Recurrent Neural Network based on Multi-Kernel Learning Support Vector Machine for Transformer Fault Diagnosis”, International Journal of Advanced Computer Science and Applications, 14(1): 125-135

130. Kang M., Kim J., Kim J. M., Tan A.C.C., Kim E.Y., Choi B.K. (2015), “Reliable fault diagnosis for low-speed bearings using individually trained support vector machines with kernel discriminative feature analysis”, IEEE Transactions on Power Electronics, 30(5): 2786–2797, DOI: 10.1109/TPEL.2014.2358494

131. Zhu X., Xiong J., Liang Q. (2018), “Fault diagnosis of rotation machinery based on support vector machine optimized by quantum genetic algorithm”, IEEE Access, 6: 33583–33588, DOI: 10.1109/ACCESS.2018.2789933

132. Li X., Zheng A., Zhang X., Li C., Zhang L. (2013), “Rolling element bearing fault detection using support vector machine with improved ant colony optimization”, Measurement 46(8): 2726–2734, DOI: 10.1016/j.measurement.2013.04.081

133. Su Z., Tang B., Liu Z., Qin Y. (2015), “Multi-fault diagnosis for rotating machinery based on orthogonal supervised linear local tangent space alignment and least square support vector machine”, Neurocomputing 157: 208–222, DOI: 10.1016/j.neucom.2015.01.016

134. Li Y., Miao B., Zhang W., Chen P., Liu J., Jiang X. (2019), “Refined composite multiscale fuzzy entropy: Localized defect detection of rolling element bearing”, Journal of Mechanical Science and Technology, 33(1): 109–120, DOI: 10.1007/s12206-018-1211-8

135. Dong S., Xu X., Liu J., Gao Z. (2015), “Rotating machine fault diagnosis based on locality preserving projection and back propagation neural network-support vector machine model”, Measurement & Control 48(7): 211–216, DOI: 10.1177/0020294015595995

136. Kim G., Lee S., Kim S. (2014), “A novel hybrid intrusion detection method integrating anomaly detection with misuse detection”, Expert Systems with Applications, 41(4): 1690-1700, DOI10.1016/j.eswa.2013.08.066

137. Sun W., Chen J., Li J. (2007), “Decision tree and PCA-based fault diagnosis of rotating machinery”, Mechanical Systems and Signal Processing, 21(3): 1300–1317, DOI: 10.1016/j.ymssp.2006.06.010

138. Zhang Y., Li X., Gao L., Wang L., Wen L. (2018), “Imbalanced data fault diagnosis of rotating machinery using synthetic oversampling and feature learning”, Journal of Manufacturing Systems 48: 34–50, DOI: 10.1016/j.jmsy.2018.04.005

139. Sakthivel N.R., Sugumaran V., Babudevasenapati S. (2010), “Vibration based fault diagnosis of monoblock centrifugal pump using decision tree”, Expert Systems with Applications 37(6): 4040–4049, DOI: 10.1016/j.eswa.2009.10.002

140. Praveenkumar T., Sabhrish B., Saimurugan M., Ramachandran K.I. (2018), “Pattern recognition based on-line vibration monitoring system for fault diagnosis of automobile gearbox”, Measurement 114: 233–242, DOI: 10.1016/j.measurement.2017.09.041

141. Breiman L. (2001), “Random forests”, Machine Learning 45(1) 5–32, DOI: 10.1023/A:1010950718922

142. Li C., Sanchez R.V., Zurita G., Cerrada M., Cabrera D., Vasquez R.E. (2016), “Gearbox fault diagnosis based on deep random forest fusion of acoustic and vibratory signals”, Mechanical Systems and Signal Processing, 76–77: 283–293, DOI: 10.1016/j.ymssp.2016.02.007

143. Wang Z., Zhang Q., Xiong J., Xiao M., Sun G., He J. (2017), “Fault diagnosis of a rolling bearing using wavelet packet denoising and random forests”, IEEE Sensors Journal 17(17): 5581–5588, DOI: 10.1109/JSEN.2017.2726011

144. Tun W., Wong J. K., Ling S. (2021), “Hybrid Random Forest and Support Vector Machine Modeling for HVAC Fault Detection and Diagnosis”, Sensors, 21(24): 8163, DOI: 10.3390/s21248163

145. Asadi S., Roshan S., Kattan M. W. (2021), “Random forest swarm optimization-based for heart diseases diagnosis”, Journal of Biomedical Informatics, 115: 103690, DOI: 10.1016/j.jbi.2021.103690

146. Yu J., Bai M., Wang G., Shi X. (2018), “Fault diagnosis of planetary gearbox with incomplete information using assignment reduction and flexible naïve Bayesian classifier”, Journal of Mechanical Science and Technology, 32(1): 37–47, DOI: 10.1007/s12206-017-1205-y

147. Wang Z., Wang L., Tan Y., Yuan J., Li X. (2021), “Fault diagnosis using fused reference model and Bayesian network for building energy systems”, Journal of Building Engineering, 34: 101957, DOI: 10.1016/j.jobe.2020.101957

148. Yu J., He Y. (2018), “Planetary gearbox fault diagnosis based on data-driven valued characteristic multigranulation model with incomplete diagnostic information”, Journal of Sound and Vibration, 429: 63–77, DOI: 10.1016/j.jsv.2018.05.020

149. He Y., Wang R., Kwong S., Wang X. (2014), “Bayesian classifiers based on probability density estimation and their applications to simultaneous fault diagnosis”, Information Sciences, 259: 252–268, DOI: 10.1016/j.ins.2013.09.003

150. Yu J., Ding B., He Y. (2018), “Rolling bearing fault diagnosis based on mean multigranulation decision-theoretic rough set and non-naive Bayesian classifier”, Journal of Mechanical Science and Technology 32(11): 5201–5211, DOI: 10.1007/s12206-018-1018-7

151. Jia Y., Xu M., Wang R. (2018), “Symbolic important point perceptually and hidden Markov model based hydraulic pump fault diagnosis method”, Sensors 18(12): 4460, DOI: 10.3390/s18124460

152. Zhou H., Chen J., Dong G., Wang R. (2016), “Detection and diagnosis of bearing faults using shift-invariant dictionary learning and hidden Markov model”, Mechanical Systems and Signal Processing, 72–73: 65–79, DOI: 10.1016/j.ymssp.2015.11.022

153. Huang D., Ke L., Chu X., Zhao L., Mi B. (2018), “Fault diagnosis for the motor drive system of urban transit based on improved hidden Markov model”, Microelectronics Reliability, 82: 179–189, DOI: 10.1016/j.microrel.2018.01.017

154. Xiao W., Chen J., Dong G., Zhou Y., Wang Z. (2012), “A multichannel fusion approach based on coupled hidden markov models for rolling element bearing fault diagnosis”, Proceedings of The Institution of Mechanical Engineers Part C-Journal of Mechanical Engineering Science, 226(C1): 202–216, DOI: 10.1177/0954406211412015

155. LeCun Y., Bengio Y., Hinton G., “Deep learning”, Nature 521(7553): 436–444, DOI: 10.1038/nature14539

156. Karras T., Laine S., and Aila T. (2021), “A style-based generator architecture for generative adversarial networks”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(12): 4217-4228, DOI: 10.1109/TPAMI.2020.2970919

157. Voulodimos A., Doulamis N., Doulamis A., Protopapadakis E. (2018), “Deep Learning for Computer Vision: A Brief Review”, Computational Intelligence and Neuroscience, 2018: 7068349, DOI: 10.1155/2018/7068349

158. Schütt K. T., Gastegger M., Tkatchenko A., Müller K.-R., Maurer R. J. (2019), “Unifying machine learning and quantum chemistry with a deep neural network for molecular wavefunctions,” Nature Communications, 10(1): 5024, DOI: 10.1038/s41467-019-12875-2

159. Arcadu F., Benmansour F., Maunz A., Willis J., Haskova Z., Prunotto M. (2019), “Deep learning algorithm predicts diabetic retinopathy progression in individual patients,” NPJ Digital Medicine, 2(1): 1–9, DOI: 10.1038/s41746-019-0172-3

160. Abdelhamid A. A., El-Kenawy E. M., Alotaibi B., Amer G., Abdelkader M. Y., Ibrahim A., Eid, M. M. (2022), “Robust Speech Emotion Recognition Using CNN plus LSTM Based on Stochastic Fractal Search Optimization Algorithm”, IEEE Access, 10: 49265-49284, DOI: 10.1109/ACCESS.2022.3172954

161. Chorowski J., Weiss R. J., Bengio S., van den Oord A. (2019), “Unsupervised speech representation learning using WaveNet autoencoders,” IEEE-ACM Transactions on Audio Speech and Language Processing, 27(12): 2041–2053, DOI: 10.1109/TASLP.2019.2938863

162. S. Schneider, A. Baevski, R. Collobert, and M. Auli (2019), “wav2vec: Unsupervised pre-training for speech recognition”, Interspeech 2019, 3465–3469, DOI: 10.21437/Interspeech.2019-1873

163. Khurana D., Koli A., Khatter K., Singh S. (2023), “Natural language processing: state of the art, current trends and challenges”, Multimedia Tools and Applications, 82(3): 3713-3744, DOI: 10.1007/s11042-022-13428-4

164. Peters M., Neumann M., Iyyer M., Gardner M. (2018), “Deep contextualized word representations,” North American Chapter of the Association for Computational Linguistics, 2018, pp. 2227–2237, DOI: 10.18653/v1/N18-1202

165. Song Y., Hyun S., Cheong Y.G. (2021), “Analysis of Autoencoders for Network Intrusion Detection”, Sensors 21(13): 4294, DOI: 10.3390/s21134294

166. Jia F., Lei Y., Lin J., Zhou X., Lu N. (2016), “Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data”, Mechanical Systems and Signal Processing, 72–73: 303–315, DOI10.1016/j.ymssp.2015.10.025

167. Lu C., Wang Z., Qin W., Ma J. (2017), “Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification”, Signal Processing, 130: 377–388, DOI: 10.1016/j.sigpro.2016.07.028

168. Dou J., Xu C., Jiao S., Li B., Zhang J., Xu X. (2020), “An unsupervised online monitoring method for tool wear using a sparse auto-encoder”, International Journal of Advanced Manufacturing Technology 106(5): 2493–2507, DOI:10.1007/s00170-019-04788-7

169. Ochoa L. E. E., Quinde I. B. R., Sumba J. P. C., Guevara A. V. Jr, Morales-Menendez R., (2019) “New approach based on autoencoders to monitor the tool wear condition in HSM”. IFAC PapersOnLine, 52(11):206–211, DOI: 10.1016/j.ifacol.2019.09.142

170. Kim J., Lee H., Jeon J. W., Kim J. M., Lee H. U., Kim S. (2020) “Stacked auto-encoder based CNC tool diagnosis using discrete wavelet transform feature extraction”. Processes 8(4):456, DOI: 10.3390/pr8040456

171. Liu H., Zhou J., Zheng Y., Jiang W., Zhang Y. (2018), “Fault diagnosis of rolling bearings with recurrent neural network based autoencoders”, ISA Transactions, 77: 167–178, DOI: 10.1016/j.isatra.2018.04.005

172. Shi M., Ding C., Wang R., Song Q., Shen C., Huang W. Zhu Z. (2022) “Deep hypergraph autoencoder embedding: An efficient intelligent approach for rotating machinery fault diagnosis”, Knowledge-Based Systems, 260: 110172, DOI: 10.1016/j.knosys.2022.110172

173. Mao W., Feng W., Liang X. (2019), “A novel deep output kernel learning method for bearing fault structural diagnosis”, Mech. Syst. Signal Process. 117: 293–318

174. Liu X., Zhou Q., Zhao J., Shen H., Xiong X. (2019), “Fault diagnosis of rotating machinery under noisy environment conditions based on a 1-D convolutional autoencoder and 1-D convolutional neural network”, Sensors 19(4): 972, DOI: 10.3390/s19040972

175. Li J., Li X., He D., Qu Y. (2019), “A novel method for early gear pitting fault diagnosis using stacked SAE and GBRBM”, Sensors 19(4): 758, DOI: 10.3390/s19040758

176. Chen Z., Li W. (2017), “Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network”, IEEE Transactions on Instrumentation and Measurements 66(7): 1693–1702, DOI: 10.1109/TIM.2017.2669947

177. Zhang N., Ding S., Zhang J., Xue Y. (2018), “An overview on restricted Boltzmann machines”. Neurocomputing 275: 1186–1199, DOI: 10.1016/j.neucom.2017.09.065

178. Fu Y., Zhang Y., Qiao H., Li D., Zhou H., Leopold J. (2015), “Analysis of feature extracting ability for cutting state monitoring using deep belief networks”, 15th CIRP Conference on Modelling of Machining Operations (15TH CMMO), 31: 29–34, DOI: 10.1016/j.procir.2015.03.016

179. Chen Y., Jin Y., Jiri G. (2018), “Predicting tool wear with multi-sensor data using deep belief networks”, International Journal of Advanced Manufacturing Technology, 99(5-8): 1917–1926, DOI: 10.1007/s00170-018-2571-z

180. Chu Y., Zhao X., Zou Y., Xu W., Han J., Zhao Y. (2018), “A Decoding Scheme for Incomplete Motor Imagery EEG With Deep Belief Network”, Frontiers in Neuroscience, 12: 680, DOI: 10.3389/fnins.2018.00680

181. Tamilselvan P., Wang P.F. (2013), “Failure diagnosis using deep belief learning based health state classification”, Reliability Engineering & System Safety, 115: 124–135, DOI: 10.1016/j.ress.2013.02.022

182. Jiang H., Shao H., Chen X., Huang J. (2018), “A feature fusion deep belief network method for intelligent fault diagnosis of rotating machinery”, Journal of Intelligent & Fuzzy Systems, 34(6): 3513–3521, DOI: 10.3233/JIFS-169530

183. He X., Wang D., Li Y., Zhou C. (2016), “A novel bearing fault diagnosis method based on Gaussian restricted boltzmann machine”, Mathematical Problems in Engineering, 2016: 2957083, DOI: 10.1155/2016/2957083

184. Zhu J., Hu T., Jiang B., Yang X. (2020), “Intelligent bearing fault diagnosis using PCA-DBN framework”, Neural Computing & Applications, 32(14): 10773-10781, DOI: 10.1007/s00521-019-04612-z

185. Yu D., Chen Z., Xiahou K., Li M., Ji T., Wu Q. (2018), “A radically data-driven method for fault detection and diagnosis in wind turbines”, International Journal of Electrical Power & Energy Systems, 99: 577–584, DOI: 10.1016/j.ijepes.2018.01.009

186. Wang X., Huang J., Ren G., Wang D. (2017), “A hydraulic fault diagnosis method based on sliding-window spectrum feature and deep belief network”, Journal of Vibroengineering 19(6): 4272–4284, DOI: 10.21595/jve.2017.18549

187. Shao S., Sun W., Yan R., Wang P., Gao R. X. (2017), “A deep learning approach for fault diagnosis of induction motors in manufacturing”, Chinese Journal of Mechanical Engineering, 30(6): 1347–1356, DOI: 10.1007/s10033-017-0189-y

188. Niu G. X., Wang X., Golda M., Mastro S., Zhang B. (2021), “An optimized adaptive PReLU-DBN for rolling element bearing fault diagnosis”, Neurocomputing, 445: 26-34, DOI: 10.1016/j.neucom.2021.02.078

189. Shao H., Jiang H., Wang F., Wang Y. (2017), “Rolling bearing fault diagnosis using adaptive deep belief network with dual-tree complex wavelet packet”, ISA Transactions 69: 187–201, DOI: 10.1016/j.isatra.2017.03.017

190. He J., Yang S., Gan C. (2017), “Unsupervised fault diagnosis of a gear transmission chain using a deep belief network”, Sensors 17(7): 1564, DOI: 10.3390/s17071564

191. Jin Z., He D., Wei Z. (2022), “Intelligent fault diagnosis of train axle box bearing based on parameter optimization VMD and improved DBN”, Engineering Applications of Artificial Intelligence, 110: 104713, DOI: 10.1016/j.engappai.2022.104713

192. Krizhevsky A., Sutskever I., Hinton G.E. (2017), “Imagenet classification with deep convolutional neural networks”, Communications of the ACM 60(6): 84–90, DOI: 10.1145/3065386

193. Gu J., Wang Z., Kuen J., Ma L., Shahroudy A., Shuai B., Liu T., Wang X., Wang G., Cai J., Chen T. (2018), “Recent advances in convolutional neural networks”, Pattern Recognition, 77: 354–377, DOI: 10.1016/j.patcog.2017.10.013

194. Yao Y., Wang H., Li S., Liu Z., Gui G., Dan Y., Hu J. (2018), “End-to-end convolutional neural network model for gear fault diagnosis based on sound signals”, Applied Sciences-Basel, 8 (9): 1584, DOI: 10.3390/app8091584

195. Li X., Li J., Qu Y., He D. (2019), “Gear pitting fault diagnosis using integrated CNN and GRU network with both vibration and acoustic emission signals”, Applied Sciences-Basel, 9(4): 768, DOI: 10.3390/app9040768

196. Emmanuel S., Yihun Y., Ahmedabadi Z. N., Boldsaikhan E. (2021), “Planetary gear train microcrack detection using vibration data and convolutional neural networks”, Neural Computing & Applications, 33(24): 17223-17243, DOI: 10.1007/s00521-021-06314-x

197. Appana D.K., Prosvirin A., Kim J.M. (2018), “Reliable fault diagnosis of bearings with varying rotational speeds using envelope spectrum and convolution neural networks”, Soft Computing, 22(20): 6719–6729, DOI: 10.1007/s00500-018-3256-0

198. Shao X., Kim C. (2022) “Unsupervised Domain Adaptive 1D-CNN for Fault Diagnosis of Bearing”, Sensors, 22(11): 4156, DOI: 10.3390/s22114156

199. Ince T., Kiranyaz S., Eren L., Askar M., Gabbouj M. (2016), “Real-time motor fault detection by 1-D convolutional neural networks”, IEEE Trans. Ind. Electron. 63 (2016) 7067–7075.

200. Gu Y., Zhang Y., Yang M., Li C. (2023), “Motor On-Line Fault Diagnosis Method Research Based on 1D-CNN and Multi-Sensor Information”, Applied Sciences-Basel, 13(7): 4192, DOI: 10.3390/app1307419

201. Janssens O., Van de Walle R., Loccufier M., Van Hoecke S. (2018), “Deep learning for infrared thermal image based machine health monitoring”, IEEE-ASME Transactions on Mechatronics, 23(1): 151–159, DOI: 10.1109/TMECH.2017.2722479

202. Wen L., Li X., Gao L., Zhang Y. (2018), “A new convolutional neural network-based data-driven fault diagnosis method”, IEEE Transactions on Industrial Electronics, 65(7): 5990–5998, DOI: 10.1109/TIE.2017.2774777

203. Islam M. M. M., Kim J. M. (2019), “Automated bearing fault diagnosis scheme using 2D representation of wavelet packet transform and deep convolutional neural network”, Computers in Industry 106: 142–153, DOI: 10.1016/j.compind.2019.01.008

204. Huang D., Zhang W., Guo F., Liu W., Shi X. (2023), “Wavelet Packet Decomposition-Based Multiscale CNN for Fault Diagnosis of Wind Turbine Gearbox”, IEEE Transactions on Cybernetics, 53(1): 443-453, DOI: 10.1109/TCYB.2021.3123667

205. Guo S., Yang T., Gao W., Zhang C., Zhang Y. (2018), “An intelligent fault diagnosis method for bearings with variable rotating speed based on pythagorean spatial pyramid pooling CNN”, Sensors, 18(11): 3857, DOI: 10.3390/s18113857

206. Zhang J., Zhou Y., Wang B., Wu Z. (2021), “Bearing fault diagnosis base on multi-scale 2D-CNN model”, 2021 3rd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI 2021), 72-75, DOI: 10.1109/MLBDBI54094.2021.00021

207. Cao X., Chen B., Yao B., He W. (2019), “Combining translation-invariant wavelet frames and convolutional neural network for intelligent tool wear state identification”, Computers in Industry, 106: 71–84, DOI: 10.1016/j.compind.2018.12.018

208. Kumar A., Gandhi C. P., Zhou Y., Vashishtha G., Kumar R., Xiang J. (2020), “Improved CNN for the diagnosis of engine defects of 2-wheeler vehicle using wavelet synchro-squeezed transform (WSST)”, Knowledge-Based Systems, 208: 106453, DOI: 10.1016/j.knosys.2020.106453

209. Xia M., Li T., Xu L., Liu L., de Silva C.W. (2018), “Fault diagnosis for rotating machinery using multiple sensors and convolutional neural networks”, IEEE-ASME Transactions on Mechatronics, 23(1): 101–110, DOI: 10.1109/TMECH.2017.2728371

210. Yu Y., Si X., Hu C., Zhang J. (2019), “A review of recurrent neural networks: LSTM cells and network architectures”. Neural Computation, 31(7): 1235–1270, DOI: 10.1162/neco_a_01199

211. Houdt G. V., Mosquera C., Napoles G. (2020), “A review on the long short-term memory model “, Artificial Intelligence Review 53(8): 5929–5955, DOI: 10.1007/s10462-020-09838-1

212. Vashisht R. K., Peng Q. (2021), “Online chatter detection for milling operations using LSTM neural networks assisted by motor current signals of ball screw drives”. Journal Of Manufacturing Science and Engineering-Transactions of the ASME 143(1): 011008, DOI: 10.1115/1.4048001

213. Zou P., Hou B., Jiang L., Zhang Z, (2020), “Bearing Fault Diagnosis Method Based on EEMD and LSTM”, International Journal of Computers Communications & Control, 15(1): 1010, DOI: 10.15837/ijccc.2020.1.3780

214. Zhao X., Alqatawneh I., Lane M., Li H., Qin Y., Gu F., Ball A. D. (2023), “Investigation into LSTM Deep Learning for Induction Motor Fault Diagnosis”, Proceedings of Income-Vi and Tepen 2021: Performance Engineering and Maintenance Engineering, 117: 505-518, DOI: 10.1007/978-3-030-99075-6_41

215. Zhao H., Sun S., Jin B. (2018), “Sequential Fault Diagnosis Based on LSTM Neural Network”, IEEE Access, 6: 12929-12939, DOI: 10.1109/ACCESS.2018.2794765

216. Lei J., Liu C., Jiang D. (2019), “Fault diagnosis of wind turbine based on Long Short-term memory networks”, Renewable Energy, 133: 422-432, DOI: 10.1016/j.renene.2018.10.031

217. Xiang L., Wang P., Yang X., Hu A., Su H. (2021), “Fault detection of wind turbine based on SCADA data analysis using CNN and LSTM with attention mechanism”, Measurement, 175: 109094, DOI: 10.1016/j.measurement.2021.109094

218. Chen X., Zhang B., Gao D. (2020) “Bearing fault diagnosis base on multi-scale CNN and LSTM model”, Journal of Intelligent Manufacturing, 32(4): 971-987, DOI: 10.1007/s10845-020-01600-2

219. Vos K., Peng Z., Jenkins C., Shahriar M. R., Borghesani P., Wang W. (2022), “Vibration-based anomaly detection using LSTM/SVM approaches”, Mechanical Systems and Signal Processing, 169: 108752, DOI: 10.1016/j.ymssp.2021.108752

220. Liu H., Zhou J., Zheng Y., Jiang W., Zhang Y. (2018), “Fault diagnosis of rolling bearings with recurrent neural network based autoencoders”, ISA Transactions, 77: 167-178, DOI: 10.1016/j.isatra.2018.04.005

221. He K., Zhang X., Ren S., Sun J. (2016), “Deep residual learning for image recognition”, in: IEEE Conference on Computer Vision and Pattern Recognition, Seattle, 2016: 770–778, DOI: 10.1109/CVPR.2016.90

222. Ma S., Chu F., Han Q. (2019), “Deep residual learning with demodulated time-frequency features for fault diagnosis of planetary gearbox under nonstationary running conditions”, Mechanical Systems and Signal Processing 127: 190–201, DOI: 10.1016/j.ymssp.2019.02.055

223. Zhang W., Li X., Ding Q. (2019), “Deep residual learning-based fault diagnosis method for rotating machinery”, ISA Transactions 95: 295–305, DOI: 10.1016/j.isatra.2018.12.025

224. Hou S., Lian A., Chu Y. (2023), “Bearing fault diagnosis method using the joint feature extraction of Transformer and ResNet”, Measurement Science and Technology, 34(7): 075108, DOI: 10.1088/1361-6501/acc885

225. Liang P., Wang W., Yuan X., Liu S., Zhang, L., Cheng Y. (2022), “Intelligent fault diagnosis of rolling bearing based on wavelet transform and improved ResNet under noisy labels and environment”, Engineering Applications of Artificial Intelligence, 115: 105269, DOI: 10.1016/j.engappai.2022.105269

226. Zhao M., Kang M., Tang B., Pecht M. (2018), “Deep residual networks with dynamically weighted wavelet coefficients for fault diagnosis of planetary gearboxes”, IEEE Transactions on Industrial Electronics 65(5): 4290–4300, DOI: 10.1109/TIE.2017.2762639

227. Peng D., Liu Z., Wang H., Qin Y., Jia L. (2019), “A novel deeper one-dimensional CNN with residual learning for fault diagnosis of wheelset bearings in highspeed trains”, IEEE Access 7: 10278–10293, DOI: 10.1109/ACCESS.2018.2888842

228. Su L., Ma L., Qin N., Huang D., Kemp A., (2019), “Fault diagnosis of high-speed train bogie by residual-squeeze net”, IEEE Transactions on Industrial Informatics, 15(7): 3856–3863, DOI: 10.1109/TII.2019.2907373

229. Hotelling H. (1933), “Analysis of a complex of statistical variables into principal components”, Journal of Educational Psychology,24: 417–441, DOI: 10.1037/h0071325

230. Hoffmann H. (2007), “Kernel PCA for novelty detection”, Pattern Recognition, 40(3): 863–874, DOI: 10.1016/j.patcog.2006.07.009

231. Schölkopf B., Platt J. C., Shawe-Taylor J., Smola A. J., Williamson R. C. (2001), “Estimating the support of a high-dimensional distribution”, Neural Computation, 13(7): 1443–1471, DOI: 10.1162/089976601750264965

232. Tax D. M. J., Duin R. P. W. (2004), “Support vector data description”, Machine Learning., 54(1): 45–66, DOI: 10.1023/B:MACH.0000008084.60811.49

233. Knorr E. M., Ng R. T., Tucakov V. (2000), “Distance-based outliers: Algorithms and applications”, VLDB Journal, 8(3-4): 237–253, DOI: 10.1007/s007780050006

234. Principi E., Vesperini F., Squartini S., Piazza F., “Acoustic novelty detection with adversarial autoencoders”, 2017 International Joint Conference on Neural Networks (IJCNN)., 2017: 3324–3330, DOI: 10.1109/IJCNN.2017.7966273

235. Perera P., Patel V. M. (2019), “Learning deep features for one-class classification”, IEEE Transactions on Image Processing, 28(11):5450–5463, DOI10.1109/TIP.2019.2917862

236. Erfani S. M., Rajasegarar S., Karunasekera S., Leckie C. (2016), “High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning,” Pattern Recognition, 58: 121–134, DOI: 10.1016/j.patcog.2016.03.028

237. Xia X., Pan X., Li N., He X., Ma L. (2022), “GAN-based anomaly detection: A review”, Neurocomputing, 493: 497-535, DOI: 10.1016/j.neucom.2021.12.093

238. Malaiya R. K., Kwon D., Kim J., Suh S. C., Kim H., Kim I. (2019), “An empirical evaluation of deep learning for network anomaly detection,” 2018 International Conference on Computing, Networking and Communications (ICNC), 2018: 893–898, DOI: 10.1109/ICCNC.2018.8390278

239. Hasan M., Islam M. M., Zarif M. I. I., Hashem M. M. A (2019), “Attack and anomaly detection in IoT sensors in IoT sites using machine learning approaches”, Internet of Things, 7: 100059, DOI: 10.1016/j.iot.2019.100059

240. Chekanov S., Hopkins W. (2022) “Event-Based Anomaly Detection for Searches for New Physics”, Universe, 8(10): 494, DOI: 10.3390/universe8100494

241. Kharkov Y. A., Sotskov V. E., Karazeev A. A., Kiktenko E. O., Fedorov K. (2020), “Revealing quantum chaos with machine learning,” Physical Review B, 101(6): 064406, DOI: 10.1103/PhysRevB.101.064406

242. Zhang J., Xie Y., Pang G., Liao Z., Verjans J., Li W., Sun Z., He J., Li Y., Shen C., Xia Y. (2021), “Viral Pneumonia Screening on Chest X-Rays Using Confidence-Aware Anomaly Detection”, IEEE Transactions on Medical Imaging, 40(3): 879-890, DOI: 10.1109/TMI.2020.3040950

243. Shvetsova N., Bakker B., Fedulova I., Schulz H., Dylov D. V. (2021), “Anomaly Detection in Medical Imaging with Deep Perceptual Autoencoders”, IEEE Access, 9: 118571-118583, DOI: 10.1109/ACCESS.2021.3107163

244. Mosavi A., Faghan Y., Ghamisi P., Duan P., Ardabili S. F., Salwana E., Band S. S. (2020), “Comprehensive Review of Deep Reinforcement Learning Methods and Applications in Economics”, Mathematics, 8(10): DOI: 10.3390/math8101640

245. Sipple J. (2020), “Interpretable, multidimensional, multimodal anomaly detection with negative sampling for detection of device failure,” in Proc. Int. Conf. Mach. Learn., 2020, pp. 4368–4377, DOI: 10.48550/arXiv.2007.10088

246. Yan S., Shao H., Xiao Y., Liu B., Wan J. (2023) “Hybrid robust convolutional autoencoder for unsupervised anomaly detection of machine tools under noises”, Robotics and Computer-Integrated Manufacturing, 79: 102441, DOI: 10.1016/j.rcim.2022.102441

247. Pang G., Shen C., Cao L., Van den Hengel A. (2021), “Deep Learning for Anomaly Detection: A Review”, 54(2): 38, DOI: 10.1145/3439950

248. Ruff L., Kauffmann J. R., Vandermeulen R. A., Montavon G., Samek W., Kloft M., Dietterich T. G., Müller K. R. (2021), “A Unifying Review of Deep and Shallow Anomaly Detection”, Proceedings of the IEEE, 109(5): 756-795, DOI: 10.1109/JPROC.2021.3052449

249. Kingma D. P., Welling M. (2014), “Auto-encoding variational Bayes”, International Conference on Learning Representations, 2013, DOI: 10.48550/arXiv.1312.6114

250. Vincent P., Larochelle H., Lajoie I., Bengio Y., Manzagol P. (2010), “Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion”, Journal of Machine Learning Research, 11: 3371–3408

251. Xiong Y., Zuo R. (2022), “Robust Feature Extraction for Geochemical Anomaly Recognition Using a Stacked Convolutional Denoising Autoencoder”, Matchematical Geosciences, 54(3): 623-644, DOI: 10.1007/s11004-021-09935-z

252. Makhzani A., Frey B. (20, “k-sparse autoencoders,” International Conference on Learning Representations (ICLR), 2014, DOI: 10.48550/arXiv.1312.5663

253. Arpit D., Zhou Y., Ngo H., Govindaraju V. (2016) “Why regularized auto-encoders learn sparse representation?”, International Conference on Machine Learning, 48: 136–144

254. Zhou C., Paffenroth R. C. (2017), “Anomaly detection with robust deep autoencoders”, Knowledge Discovery and Data Mining, 2017: 665–674, DOI:10.1145/3097983.3098052

255. Rifai S., Vincent P., Muller X., Glorot X., Bengio Y. (2011), “Contractive auto-encoders: Explicit invariance during feature extraction”, International Conference on Machine Learning, 2011: 833–840

256. Chen J., Sathe S., Aggarwal C. C., Turaga D. S. (2017), “Outlier detection with autoencoder ensembles”, in Proc. SIAM Int. Conf. Data Mining, 2017: 90–98, DOI: 10.1137/1.9781611974973.11

257. Sarafijanovic-Djukic N., Davis J. (2019), “Fast distance-based anomaly detection in images using an inception-like autoencoder”, Discovery Science, 2019: 493–508, DOI: 10.1007/978-3-030-33778-0_37

258. T. Amarbayasgalan, B. Jargalsaikhan, and K. Ryu (2018), “Unsupervised novelty detection using deep autoencoders with density based clustering”, Applied Sciences-Basel, 8(9): 1468, DOI: 10.3390/app8091468

259. Arjovsky M., Chintala S., Bottou L. (2017), “Wasserstein generative adversarial networks”, 34th International Conference on Machine Learning, 70, 214–223, DOI: 10.48550/arXiv.1701.07875

260. Li Z., Zheng T., Wang Y., Cao Z., Guo Z., Fu, H. (2021) “A Novel Method for Imbalanced Fault Diagnosis of Rotating Machinery Based on Generative Adversarial Networks”, IEEE Transactions on Instrumentation and Measurement, 70: 3500417, DOI: 10.1109/TIM.2020.3009343

261. Schlegl T., Seeböck P., Waldstein S. M., Schmidt-Erfurth U., Langs G. (2017), “Unsupervised anomaly detection with generative adversarial networks to guide marker discovery,” Information Processing in Medical Imaging, 10265: 146–157, DOI: 10.1007/978-3-319-59050-9_12

262. Akcay S., Atapour-Abarghouei A., Breckon T.P. (2018), “Ganomaly: Semi-supervised anomaly detection via adversarial training”, Asian conference on computer vision, 11363: 622–637, DOI: 10.1007/978-3-030-20893-6_39

263. Zenati H., Foo C.S., Lecouat B., Manek G., Chandrasekhar V.R. (2018), “Efficient gan-based anomaly detection”, 20th IEEE International Conference on Data Mining (ICDM), DOI: 10.48550/arXiv.1802.06222

264. Zenati H., Romain M., Foo C-S., Lecouat B., Chandrasekhar V. (2018), “Adversarially learned anomaly detection”, 2018 IEEE International conference on data mining (ICDM), 2018: 727-736, DOI: 10.1109/ICDM.2018.00088

265. Schlegl T., Seeböck P., Waldstein S. M., Langs G., Schmidt-Erfurth U. (2019), “f-anogan: Fast unsupervised anomaly detection with generative adversarial networks”, Medical image analysis, 54: 30-44, DOI: 10.1016/j.media.2019.01.010

266. An J., Cho S. (2015), “Variational autoencoder based anomaly detection using reconstruction probability,” Special Lecture on IE, 2(1): 1–18

267. Yao R., Liu C., Zhang L., Peng, P. (2019), “Unsupervised Anomaly Detection Using Variational Auto-Encoder based Feature Extraction”, 2019 IEEE International Conference on Prognostics and Health Management (ICPHM)

268. Xu H., Chen W., Zhao N., Li Z., Bu J., Li, Z., Liu Y., Zhao Y., Pei D., Feng Y. (2018), “Unsupervised anomaly detection via variational auto-encoder for seasonal KPIs in Web applications,” Web Conference 2018: Proceedings of The World Wide Web Conference (WWW2018)., 2018: 187–196, DOI: 10.1145/3178876.3185996

269. Nalisnick E., Matsukawa A., Teh Y. W., Gorur D., Lakshminarayanan B. (2019), “Do deep generative models know what they don’t know?” International Conference on Learning Representations, DOI: 10.48550/arXiv.1810.09136

270. Lin S., Clarke R., Birke R., Schönborn S., Trigoni N., Roberts, S. (2020), “ANOMALY DETECTION FOR TIME SERIES USING VAE-LSTM HYBRID MODEL”, International Conference on Acoustics Speech and Signal Processing ICASSP, 2020: 4322-4326, DOI: 10.1109/icassp40776.2020.9053558

271. Park D., Hoshi Y., Kemp C. C. (2018), “A multimodal anomaly detector for robot-assisted feeding using an LSTM-based variational autoencoder,” IEEE Robotics and Automation Letters, 3(3): 1544–1551, DOI: 10.1109/LRA.2018.2801475

272. Luo P., Wang B., Li T., Tian J. (2021), “ADS -B anomaly data detection model based on VAE-SVDD”, Computers & Security, 104: 102213, DOI: 10.1016/j.cose.2021.102213

273. Zhou Y., Liang X., Zhang W., Zhang L., Song X. (2021), “VAE-based Deep SVDD for anomaly detection”, Neurocomputing, 453: 131-140, DOI: 10.1016/j.neucom.2021.04.089

274. Jain A. K. (2010), “Data clustering: 50 years beyond K-means”, Pattern Recognition Letters, 31(8): 651–666, DOI: 10.1016/j.patrec.2009.09.011

275. Wang Z., Zhou Y., Li G. (2020), “Anomaly Detection by Using Streaming K-Means and Batch K-Means”, 2020 5th IEEE International Conference on Big Data Analytics (IEEE ICBDA 2020), 2020: 11-17

276. Chang C., Hsu W., Liao I. (2019), “Anomaly Detection for Industrial Control Systems Using K-Means and Convolutional Autoencoder”, 2019 27th International Conference on Software, Telecommunications and Computer Networks (SOFTCOM), 2019: 136-141, DOI: 10.23919/softcom.2019.8903886

277. Dhillon I. S., Guan Y., Kulis B. (2004), “Kernel k-means, spectral clustering and normalized cuts,” Knowledge Discovery Data Mining, 2004: 551–556, DOI: 10.1145/1014052.1014118

278. Görnitz N., Lima L. A., Müller K.-R., Kloft M., Nakajima S. (2018), “Support vector data descriptions and k-means clustering: One class?”, IEEE Transactions on Neural Networks and Learning Systems, 29(9): 3994–4006, Sep. 2018, DOI: 10.1109/TNNLS.2017.2737941

279. Amruthnath N., Gupta T. (2018), “A research study on unsupervised machine learning algorithms for early fault detection in predictive maintenance”, 2018 5th International Conference on Industrial Engineering and Applications (ICIEA), 2018: 355–361

280. Liu F. T., Ting K. M., Zhou Z. (2008) “Isolation Forest”, IEEE International Conference on Data Mining (ICDM), 2008, DOI: 10.1109/ICDM.2008.17

281. M. Kampffmeyer, S. Løkse, F. M. Bianchi, L. Livi, A.-B. Salberg, and R. Jenssen, “Deep divergence-based approach to clustering,” Neural Netw., vol. 113, pp. 91–101, May 2019

282. Xie J., Girshick R., Farhadi A. (2016), “Unsupervised deep embedding for clustering analysis,” International Conference on Machine Learning, 48: 478–487

283. An P., Wang Z., Zhang C. (2022), “Ensemble unsupervised autoencoders and Gaussian mixture model for cyberattack detection”, Information Processing & Management, 59(2): 102844, DOI: 10.1016/j.ipm.2021.102844

284. Obied M. A., Ghaleb F. F. M., Hassanien A. E., Abdelfattah A. M. H., Zakaria W. (2023) “Deep Clustering-Based Anomaly Detection and Health Monitoring for Satellite Telemetry”, Big Data and Cognitive Computing, 7(1): 39, DOI: 10.3390/bdcc7010039

285. Ruff L., Vandermeulen R. A., Görnitz N., Deecke L., Siddiqui S. A., Binder A., Müller E., Kloft M. (2018), “Deep one-class classification,” International Conference on Machine Learning, 80: 4390–4399

286. Jolliffe I. T. (2002), “Principal Component Analysis”, (Springer Series in Statistics), Springer, ISBN: 0387954422

287. Tipping M. E., Bishop C. M. (1999), “Probabilistic principal component analysis”, Journal of The Royal Statistical Society Series B-Statistical Methodology, 61(3): 611–622, DOI: 10.1111/1467-9868.00196

288. Sharan V., Gopalan P., Wieder U. (2018), “Efficient anomaly detection via matrix sketching”, Advances in Neural Information Processing Systems (NIPS 2018)., 31: 8069-8080

289. O’Reilly C., Gluhak A., Imran M. A. (2016), “Distributed Anomaly Detection Using Minimum Volume Elliptical Principal Component Analysis”, IEEE Transactions on Knowledge and Data Engineering, 28(9): 2320-2333, DOI: 10.1109/TKDE.2016.2555804

290. Shyu M.-L., Chen S.-C., Sarinnapakorn K., Chang L. (2006), “A novel anomaly detection scheme based on principal component classifier”, Foundations and Novel Approaches in Data Mining, 9: 311-+

291. Khan S. S., Madden M. G. (2014), “One-class classification: Taxonomy of study and review of techniques”, Knowledge Engineering Review, 29(3): 345–374, DOI10.1017/S026988891300043X

292. Menon A. K., Williamson R. C. (2018), “A loss framework for calibrated anomaly detection,” Advances in Neural Information Processing Systems (NIPS 2018) 31: 1494-1504

293. Lee G., Scott C. (2010), “Nested support vector machines,” IEEE Transactions on Signal Processing, 58(3): 1648–1660, DOI: 10.1109/TSP.2009.2036071

294. Gautam C., Balaji R., Sudharsan K., Tiwari A., Ahuja K. (2019), “Localized multiple kernel learning for anomaly detection: One-class classification”, Knowledge-Based Systems, 241–252, DOI10.1016/j.knosys.2018.11.030

295. Stolpe M., Bhaduri K., Das K., Morik K. (2013), “Anomaly detection in vertically partitioned data by distributed Core vector machines”, Joint European Conference on Machine Learning and Knowledge Discovery in Databases, 2013: 321–336, DOI: 10.1007/978-3-642-40994-3_21

296. Ghasemi A., Rabiee H. R., Manzuri M. T., Rohban M. H. (2012), “A Bayesian approach to the data description problem”, AAAI Conference of Artifical Intelligence, 2(1): 907–913, DOI:10.1609/aaai.v26i1.8288

297. Anaissi A., Khoa N. L. D., Rakotoarivelo T., Alamdari M. M., Wang Y. (2018), “Adaptive Online One-Class Support Vector Machines with Applications in Structural Health Monitoring”, ACM Transactions on Intelligent Systems and Technology, 9(6): 64, DOI: 10.1145/3230708

298. Xiao Y., Wang H., Xu W., Zhou J. (2016), “Robust one-class SVM for fault detection”, Chemometrics and Intelligent Laboratory Systems, 151: 15-25, DOI: 10.1016/j.chemolab.2015.11.010

299. Kim M., Kim J., Yu J., Choi J. K. (2023), “Active anomaly detection based on deep one-class classification”, Pattern Recognition Letters, 167: 18-24, DOI: 10.1016/j.patrec.2022.12.009

300. Ghafoori Z., Leckie C. (2020), “Deep multi-sphere support vector data description”, Proceedings of the 2020 Siam International Conference on Data Mining (SDM), 2020: 109–117, DOI: 10.1137/1.9781611976236.13

301. Arellano-Espitia F., Delgado-Prieto M., Gonzalez-Abreu A., Saucedo-Dorantes J. J., Osornio-Rios R. A. (2021), “Deep-Compact-Clustering Based Anomaly Detection Applied to Electromechanical Industrial Systems”, Sensors, 21(17): 5830, DOI: 10.3390/s21175830

302. Kingma D. P., Ba J. (2015), “Adam: A method for stochastic optimization”, International Conference on Learning Representations (ICLR), 2015, DOI: 10.48550/arXiv.1412.6980

303. Nanduri A., Sherry L. (2016), “ANOMALY DETECTION IN AIRCRAFT DATA USING RECURRENT NEURAL NETWORKS (RNN)”, 2016 Integrated Communications Navigation and Surveillance (ICNS), 2016: 5C2

304. Sabokrou M., Fayyaz M., Fathy M., Moayed Z., Klette R. (2018), “Deep-anomaly: Fully convolutional neural network for fast anomaly detection in crowded scenes”, Computer Vision and Image Understanding, 172: 88-97, DOI: 10.1016/j.cviu.2018.02.006

305. Ruff L., Vandermeulen R. A., Görnitz N., Binder A., Müller E., Müller K. R., Kloft M. (2020), “Deep semi-supervised anomaly detection”, International Conference on Learning Representations (ICLR), 2020, DOI: 10.48550/arXiv.1906.02694

306. Liznerski P., Ruff L., Vandermeulen R. A., Franks B. J., Kloft M., Müller K.R. (2021), “Explainable deep one-class classification”, International Conference on Learning Representations (ICLR) 2021, DOI: 10.48550/arXiv.2007.01760

307. Chong P., Ruff L., Kloft M., Binder A. (2020), “Simple and effective prevention of mode collapse in deep one-class classification”, IEEE International Joint Conference on Neural Networks (IJCNN), 2020: 1–9, DOI: 10.1109/ijcnn48605.2020.9207209

308. Hendrycks D., Mazeika M., Kadavath S., Song D. (2019), “Using self-supervised learning can improve model robustness and uncertainty”, ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 32 (NIPS 2019), 32: 15637–15648, DOAI: 10.48550/arXiv.1906.12340

309. Goyal S., Raghunathan A., Jain M., Simhadri H. V., Jain P. (2020), “DROCC: Deep robust one-class classification”, International Conference on Machine Learning, 119: 11335–11345, DOI: 10.48550/arXiv.2002.12718

310. Akima H. (1970), “A New Method of Interpolation and Smooth Curve Fitting Based on Local Procedures”, Journal of the ACM, 17(4): 589-602, DOI: 10.1145/321607.321609

311. Hyman J. (1983), “Accurate Monotonicity Preserving Cubic Interpolation”, SIAM JOURNAL ON SCIENTIFIC AND STATISTICAL COMPUTING, 4(4): 645-654, DOI: 10.1137/0904045

312. Pearson K. (1895), “Notes on regression and inheritance in the case of two parents”, Proceedings of the Royal Society of London., 58: 240–242, DOI: 10.1098/rspl.1895.0041

313. Murtagh F,, Legendre P, (2014), “Ward’s Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward’s Criterion?”, Journal of Classification, 31(3): 274-295, DOI: 10.1007/s00357-014-9161-z

314. Arthur D., Vassilvitskii S. (2007), “k-means++: The advantages of careful seeding”, Proceedings of The Eighteenth Annual Acm-Siam Symposium on Discrete Algorithms, 2007: 1027-1035, DOI: 10.1145/1283383.1283494

315. Elkan C. (2003), “Using the Triangle Inequality to Accelerate K-Means. Proceedings”, Twentieth International Conference on Machine Learning, 2003: 147-153

316. Zeiler M. D. (2012), “Adadelta: an adaptive learning rate method”, DOI: 10.48550/arXiv.1212.5701

317. Thorndike R. L. (1953), “Who Belongs in the Family?”. Psychometrika, 18(4): 267-276, DOI: 10.1007/BF02289263. S2CID 120467216

318. Rousseeuw P. J. (1987), “SILHOUETTES - A GRAPHICAL AID TO THE INTERPRETATION AND VALIDATION OF CLUSTER-ANALYSIS”, Journal of Computational and Applied Mathematics, 20: 53-65, DOI: 10.1016/0377-0427(87)90125-7

319. Barera D. (2019), “Cross-Validation”, Encyclopedia of Bioinformatics and Computational Biology, 1: 542-545, DOI: 10.1016/B978-0-12-809633-8.20349-X

320. Kohavi R. (1995), “A study of cross-validation and bootstrap for accuracy estimation and model selection”, 14th International Joint Conference on Artificial Intelligence – 2: 1137–1143

321. Hinton G. E., Srivastava N., Krizhevsky A., Sutskever I., Salakhutdinov R. R. (2012), “Improving neural networks by preventing co-adaptation of feature detectors”, arXiv.org (2012), arXiv:1207.0580v1, DOI: 10.48550/arXiv.1207.0580

322. Wang Z., Oates T. (2015), “Imaging Time-Series to Improve Classification and Imputation”, Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI), 2015: 3939-3945

323. Cho K., van Merrienboer B., Gulcehre C., Bahdanau D., Bougares F., Schwenk F., Bengio J. (2014), “Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation”, Association for Computational Linguistics, DOI: 10.48550/arXiv.1406.1078

324. Barnett V., Lewis T. (1994), “Outliers in Statistical Data”, 3rd ed. Hoboken, NJ, USA: Wiley, ISBN: 978-0-471-93094-5