

УДК 004.032.26:004.65

Рубець Антон Вікторович, аспірант

Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

ПОБУДОВА КОМПЛЕКСНОЇ МОДЕЛІ НЕЙРОМЕРЕЖЕВОГО АНАЛІЗУ ВЕЛИКИХ ОБСЯГІВ ДАНИХ У РЕЖИМІ РЕАЛЬНОГО ЧАСУ

Рубець А.В. Побудова комплексної моделі нейромережевого аналізу великих обсягів даних у режимі реального часу. Проведено аналіз комплексної моделі нейромережевого аналізу великих обсягів даних. Показано що в її основу може бути покладено алгоритм на базі архітектури нейромережі прямого поширення. Визначено базову процедуру прогнозування та ідентифікації системних проблем, що виникають при обробці великих масивів даних у режимі реального часу. Зазначено, що для співставлення та аналізу результатів роботи нейромереж необхідно ввести такі базові критерії оцінки, як результативність, процент невдач, середньо квадратична похибка та коефіцієнт змішаної кореляції. Наголошено, що базова модель нейромережевого алгоритму може являти собою багатошаровий перцептрон, що є елементарним видом нейромережі прямого поширення, яка характеризується мінімальною ресурсоемністю, але при цьому надає можливість побудувати мережу глибинного навчання. Побудовано залежності показників результативності, середньо квадратичної похибки та коефіцієнту змішаної кореляції в залежності від режиму навчання та контролю навчання нейромережевого алгоритму.

Ключові слова: Нейромережевий алгоритм, дата майнінг, нейромережа прямого поширення, коефіцієнт змішаної кореляції, рандомізація.

Рубець А. В. Построение комплексной модели нейросетевого анализа больших объемов данных в режиме реального времени. Проведен анализ комплексной модели нейросетевого анализа больших объемов данных. Показано что в ее основу может быть положен алгоритм на базе архитектуры нейросети прямого распространения. Отмечено, что для сопоставления и анализа результатов работы нейронных сетей необходимо ввести такие базовые критерии оценки, как результативность, процент неудач, среднее квадратическая погрешность и коэффициент смешанной корреляции. Отмечено, что базовая модель нейросетевого алгоритма может представлять собой многослойный перцептрон, что является элементарным видом нейросети прямого распространения, которая характеризуется минимальной ресурсоемкостью, но при этом дает возможность построить сеть глубинного обучения. Построены зависимости показателей результативности, среднеквадратичной погрешности и коэффициента смешанной корреляции в зависимости от режима обучения и контроля обучения нейросетевого алгоритма.

Ключевые слова: нейросетевой алгоритм, дата майнинг, нейросеть прямого распространения, коэффициент смешанной корреляции, рандомизация.

Rubets Anton. Development of the ann complex model for analysis of large amounts of data in real time. A complex model of artificial neural network analysis of large amounts of data was considered. It is shown that the model can be based on an algorithm of feedforward neural network architecture. A basic procedure for predicting and identifying system problems that arise when processing large data sets in real time is defined. The criteria that must be introduced to compare and analyze the performance of the neural networks are indicated. Dependences of performance indicators, root-mean-square error and coefficient of mixed correlation depending on the training mode and control of the neural network algorithm learning were demonstrated. The cross-validation ensemble artificial neural network technique was compared to the classic artificial neural network and modern machining learning algorithms. Finally, it was presented algorithm which included the estimation of best-trained cross-validation ensemble artificial neural network, which was used for prediction of the time of failure of the system.

Key words: neural network algorithm, data mining, feedforward neural network, coefficient of mixed correlation, randomization.

Вступ

Аналіз великих об'ємів даних у режимі реального часу є найбільш важливим завданням при розробці сучасних інформаційних систем. Розвиток мережевих ресурсів зумовлює необхідність у збільшенні персоналу, що аналізує вхідний трафік і контролює стабільність стану системи. Це призводить до експоненційного зростання витрат на організацію інформаційної системи та збільшенню ризиків пов'язаних з «людським фактором». Вирішення даної проблеми полягає у використанні нейромережевих алгоритмів, що здатні до навчання та адаптації до змін і масштабування інформаційної системи.

З метою визначення ключових аспектів зазначеної проблеми було проведено аналіз останніх досліджень і публікацій, присвячених використанню нейромережевих моделей (НММ) при аналізі великих об'ємів даних у режимі реального часу. Так, у роботах [1, 2] показані загальні принципи побудови нейромережевої архітектури та навчання нейромережевих алгоритмів у режимі реального часу. Було проведено аналіз використання з цією метою нейромереж прямого поширення [3], зокрема нейромереж на основі алгоритму рою частинок (ітераційного алгоритму рою частинок і

еволюційного алгоритму рою частинок), що показали найбільш високу ефективність. Також розглянуті еволюційні алгоритми НММ [4], алгоритми на основі нечіткої логіки [5], теорії розмитих і грубих множин [6], векторних машин [7], кепстрального метода [8, 9] і НММ з використанням експертної оцінки [10].

Тим не менш, окремі дослідження не дають можливість побудувати цілісної методології та визначити ефективну стратегію створення платформи нейромережевого аналізу великих обсягів даних у режимі реального часу. Тому метою даної роботи є проведення системного аналізу результатів досліджень, порівняння різних підходів у даній галузі та визначення критеріїв функціональності комплексної нейромережевої моделі, а також побудова відповідної методологічної бази.

Виклад основного матеріалу дослідження

1. Математична модель штучної нейронної мережі

Базовим елементами нейромережевої моделі є штучні нейрони та вагові зв'язки між ними, що являють собою елементарну обчислювальну одиницю та змінну [1, 2]. Штучні нейрони поділяються на вхідні (input), приховані (hidden) та вихідні (output). Дана класифікація функціонально відповідає поділенню нейронів біологічної нейронної мережі на рецепторні, проміжні та ефекторні, відповідно (вагові коефіцієнти у рамках даної аналогії, очевидно, відповідають синапсам). При аналізі структури НММ нейрони розбиваються на один шар вхідних нейронів, один або декілька шарів прихованих нейронів та один шар вихідних нейронів. Розглянемо два проміжні шари $h_i(h-1)$, що складаються за I і J нейронів, відповідно ($i \in [1; I]$, $j \in [1; J]$), у рамках нейромережевої архітектури (рис. 1), що утримує H прихованих шарів ($h \in [1; H]$), вхідного шару, що складається з A нейронів та вихідного шару, що складається з Y нейронів ($a \in [1; A]$, $y \in [1; Y]$)

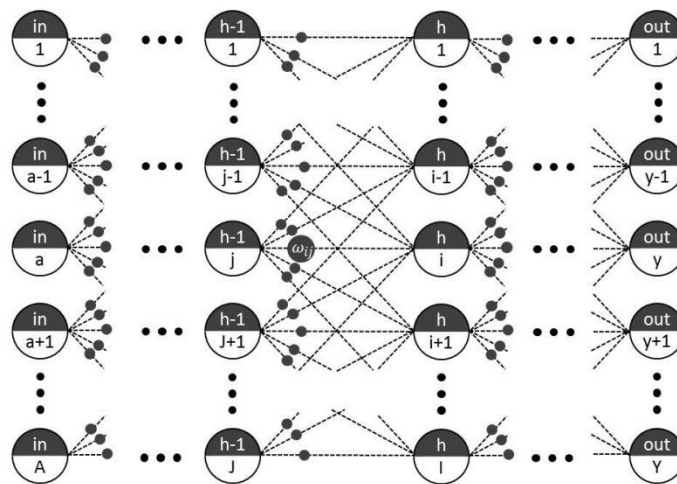


Рис. 1. Модель утворення вхідних даних нейрона у НММ прямого поширення.

При цьому вхідні дані нейрона формуються як нормалізована функцією активації (сигмоїда чи гіперболічний тангенс) сума вихідних даних нейронів попереднього шару, кожен з яких множиться на відповідні вагові коефіцієнти ω_{ij} [1, 2]. Так, для прихованого нейрона i шару h значення складатиме X_i^h :

$$\begin{cases} S_i^h = \sum_{j=1}^J \omega_{ij} \cdot X_j^{h-1} \\ \left[\begin{array}{l} X_i^h = 1 / (1 + e^{-k S_i^h}) \\ X_i^h = (e^{k S_i^h} + 1) / (e^{k S_i^h} + 1) \end{array} \right. \end{cases}, \quad (1)$$

де k — константа, а S_i^h — сума вихідних сигналів нейронів попереднього шару.

Вхідні дані нейронів вихідного шару формуються як нормалізовані суми вихідних даних нейронів останнього прихованого шару помножених на вагові коефіцієнти, вхідні дані нейронів першого шару формуються як нормалізовані суми вихідних даних нейронів вхідного шару помножених на вагові коефіцієнти, а дані нейронів вхідного шару формуються на основі нормалізованих вхідних даних всієї НММ.

Також у архітектурі НММ можуть використовуватись нейрони зміщення, що змінюють дані, які поступають на вхід інших нейронів шару шляхом зсуву функції. Вхідне та вихідне значення нейрону зміщення дорівнює «1», вони можуть бути присутні на вхідному і проміжних шарах, з'єднуючись з усіма нейронами наступного рівня, крім інших нейронів зміщення. Через те що нейрони зміщення функціонально відносять лише до обробки цифрового сигналу, вони не мають аналогів у біологічних структурах.

Загальна кількість тренувальних сетів, що проходить НММ, визначає кількість ітерацій, а кількість проходження повного набору тренувального сету визначається значенням епохи. Натренованість мережі визначається величиною помилки, що вказує на розходження між бажаною відповіддю та отриманим результатом. При правильній організації архітектури НММ з кожною новою епохою натренованість має зростати. Нехай N — кількість тренувальних сетів, Y_{ny} — вихідні дані вихідних нейронів НММ, Z_{ny} — бажані значення вихідних нейронів НММ. У такому разі, функція ефективності визначається через пошук мінімуму величини ε :

$$\varepsilon = \frac{1}{2} \sum_{n=1}^N \sum_{y=1}^Y (Y_{ny} - Z_{ny})^2 \quad (2)$$

При побудові архітектури НММ, що спеціалізується на певному класі задач, необхідно обрати гіперпараметри, до яких відносяться цільові функції роботи нейромережових алгоритмів та їх аргументи [1, 2]. До цільових функцій у даному разі зазвичай відносять:

- збіжність системи;
- швидкість навчання;
- складність організації.

У свою чергу, параметри даних функцій складаються зі значень:

- кількості прихованих шарів;
- кількість нейронів у кожному шарі;
- кількості нейронів зміщення.

Наступним етапом є попереднє навчання НММ, для чого необхідно обрати метод у навчання у відповідності за наступними класифікаціями:

- навчання зі вчителем;
- навчання без учителя;
- навчання з підкріпленням.

І, відповідно, далі виникає потреба у виборі алгоритму знаходження екстремуму цільової функції, серед яких можна виділити.

- стохастичний метод градієнтного спуску;
- пакетний метод градієнтного спуску;
- міні-пакетний метод градієнтного спуску.

Метод «навчання зі вчителем» використовується коли визначено результат, що має бути отримано на виході НММ, а відповідно і розбіжність $(Y_{ny} - Z_{ny})$. У випадку навчання з підкріпленням не існує однозначного набору Z_{ny} , але при цьому є можливість визначити динаміку збіжності системи.

Схема «навчання без учителя» використовується для виділення параметрів набору вхідних образів, та, відповідно їх класифікації за даними параметрами, тому у даному випадку не існує очевидного підходу для використання алгоритму пошуку екстремуму цільової функції збіжності системи ε .

Алгоритми градієнтного спуску, у свою чергу, розрізняються швидкістю пошуку екстремуму функції ε та ймовірності отримання локального екстремуму (що є незадовільним результатом навчання). Застосування пакетного методу полягає у обчисленні значення ε після закінчення кожної

епохи для всіх зразків, з подальшим внесенням поправок у вагові коефіцієнти кожного нейрона. При стохастичному методі поправки на вагові коефіцієнти для кожного зразка сету вводяться після обчислення вихідних даних мережі. Навчання НММ за пакетним методом характеризується перевагою по швидкості, але при цьому є вразливим по відношенню до отримання локального екстремуму цільової функції. Мініпакетний метод градієнтного спуску можна розглядати як компромісний між пакетним та стохастичним.

2. Нейромережевий аналіз великих обсягів даних

Нейромережевий аналіз даних полягає у прогнозуванні потенційних системних проблем, що базується на великій вибірці попередньо оброблених даних (рис. 2).

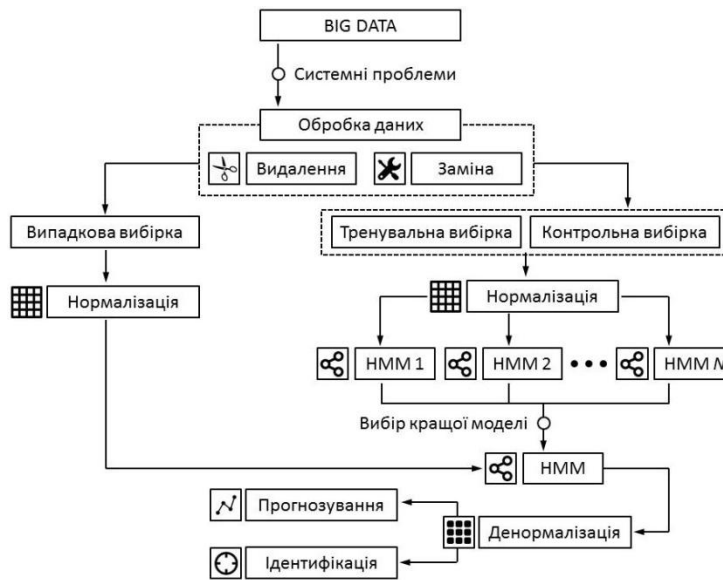


Рис. 2. Нейромережева процедура прогнозування та ідентифікації системних проблем, що виникають при обробці масивів даних типу «Big Data».

Як показано на рис. 2, нейромережевому аналізу передують аналіз стану системи та виявлення поточних проблем у режимі реального часу. Це відповідає етапу попередньої обробки даних, при якому відбувається аналіз відхилень, заміна пошкоджених даних та видалення елементів, що не становлять інформаційної цінності (як то неповних рядків таблиць, стовпців, тощо). Даний етап також може включати у себе інтеграцію, дискретизацію і форматування даних, що здійснюється з метою підвищити ефективність роботи НММ. У результаті має бути визначено ефективний обсяг повної вибірки даних, що складає основу навчальної вибірки для тренування нейромережевих алгоритмів та контрольну вибірку для оцінки їх навчання [1-3]. Ефективність навчання НММ, що може бути побудована на основі нейромережі прямого поширення (рис. 3), при цьому оцінюється у відповідності до процедури викладеної вище, тобто шляхом визначення екстремуму цільової функції.

Згідно зі схемою, представленою на рис. 3, алгоритм формування НММ базується на рандомізації фрагментів повної вибірки з метою побудови масиву навчальних вибірок та контрольної вибірки, а також рандомізації гіперпараметрів та схеми навчання нейромережевої архітектури. Базова модель нейромережевого алгоритму може являти собою багатопартийний перцептрон, що є елементарним видом нейромережі прямого поширення, яка характеризується мінімальною ресурсоемністю, але при цьому надає можливість побудувати мережу глибинного навчання [3-5]. Ефективність кожної з побудованих мереж визначається шляхом співставлення результатів застосування контрольної вибірки (збіжність системи) з кількістю тренувальних сегів (ресурсоемність моделі).

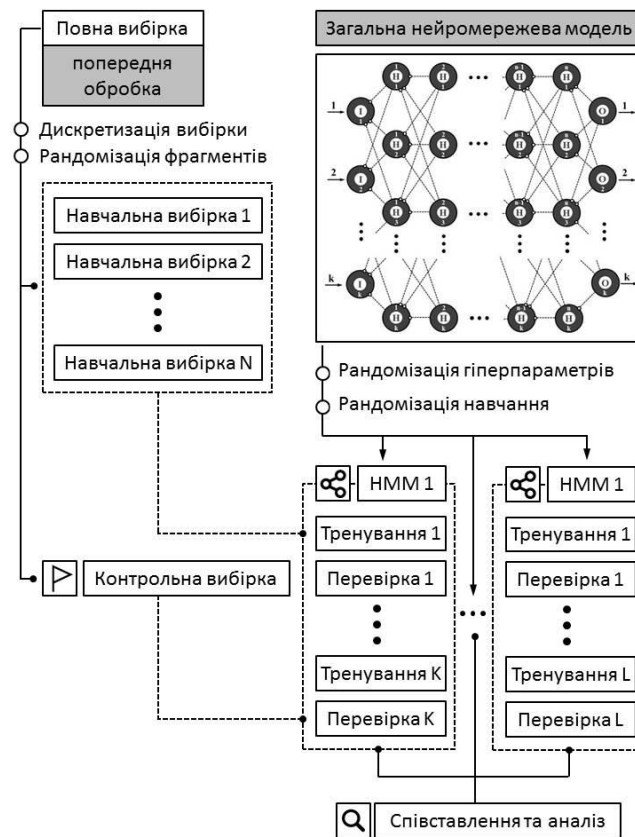


Рис. 3. Алгоритм формування архітектури НММ, що базується на основі методів рандомізації та перехресної перевірки.

3. Результати нейромережевого аналізу масивів даних у режимі реального часу

Для співставлення та аналізу результатів роботи НММ у залежності від архітектури та процедури тренування необхідно ввести базові критерії оцінки [1, 3-5], зокрема такі як результативність (HR: Hit Ratio), процент невдач (MR: Miss Ratio) та середньоквадратична похибка. Результативність k_{HR} оцінюється як співвідношення кількості вірно передбачених статусів цільових компонент навчальної вибірки N_+ до повного значення кількості даних компонент N_Σ :

$$k_{HR} = N_+ / N_\Sigma . \quad (3)$$

Аналогічним чином, процент невдач k_{MR} розраховується як величина обернена до результативності:

$$k_{MR} = 1 - N_+ / N_\Sigma = 1 - k_{HR} . \quad (4)$$

Середньоквадратична похибка σ , у свою чергу дозволяє отримати більш точний критерій оцінки. Дана величина базується на порівнянні даних (представлених у числовій формі) на виході реальної системи T_i та прогнозованої нейромережевими алгоритмами моделі даної системи P_i :

$$\sigma = \frac{\sum_{i=1}^{N_\Sigma} (T_i - P_i)^2}{N_\Sigma} . \quad (5)$$

На практиці замість величини середньоквадратичної похибки часто використовують коефіцієнт змішаної кореляції R^2 , що дозволяє визначити наскільки добре прогнозоване значення залежної змінної, отримане на основі вибраного прогнозного значення відповідає фактичним даним:

$$R^2 = \frac{(\sum_{i=1}^{N_T} ((T_i - \bar{T}_i) - (P_i - \bar{P}_i)))^2}{\sum_{i=1}^{N_T} (T_i - \bar{T}_i)^2 \cdot \sum_{i=1}^{N_P} (P_i - \bar{P}_i)^2}, \quad (6)$$

Де \bar{T}_i та \bar{P}_i відповідають середнім значенням величин T_i та P_i , відповідно.

Ключовими факторами, що визначають ефективність процесу навчання НММ у рамках розглянутої вище процедури є співвідношення етапів тренування НММ на навчальній вибірці до етапів контролю НММ на контрольній вибірці (TD-VD: Training Dataset - Validation Dataset), а також величина сукупного впливу цільових компонент (CTCI: Cumulative Target Component Influence) навчальної вибірки. CNCI визначається як співвідношення тих компонент, що потенційно впливають на стан системи і використовуються у навчальній вибірці до повної кількості компонент даного типу. Очевидно, що при збільшенні значення CNCI збільшується точність роботи НММ, але при цьому зростає час її тренування [3-5]. На рис. 4-6 показано залежність коефіцієнту змішаної кореляції (рис. 4), середньо квадратичної похибки (рис. 5) та результативності НММ (рис. 6) від значення TD-VD для різних CNCI (95,5%, 99,3%, 99,9% та 100%).

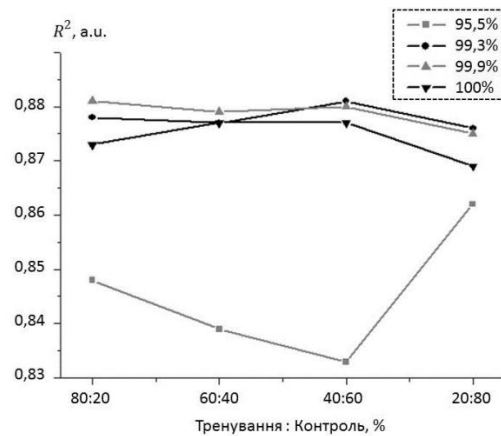


Рис. 4. Залежність коефіцієнту змішаної кореляції від співвідношення TD-VD.

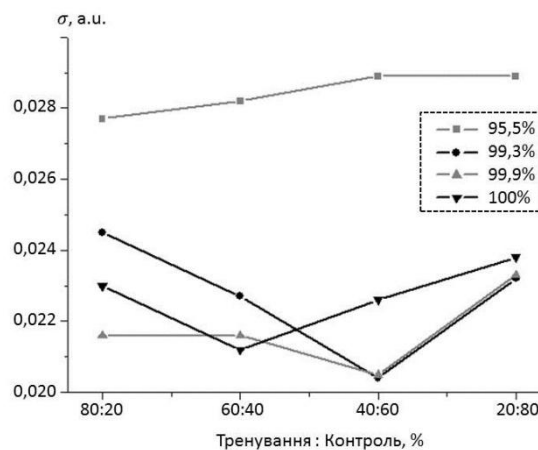


Рис. 5. Залежність середньо квадратичної похибки від співвідношення TD-VD.

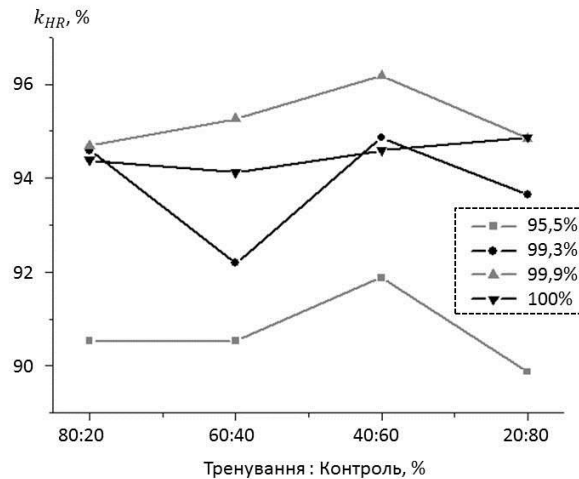


Рис. 6. Залежність результативності НММ від співвідношення TD-VD.

На основі даного алгоритму також можливо визначити оптимальну архітектуру НММ, порівнюючи різні нейромережеві алгоритми, а не режими їх роботи. Табл. 1 показує, що найбільшу ефективність при вирішенні завдання аналізу великих об'ємів даних у режимі реального часу виконують нейромережі прямого поширення, а також НММ, побудовані на їх основі [3-10].

Таблиця 1

Нейромережеві алгоритми, що демонструють найбільшу точність аналізу великих об'ємів даних інформаційних систем при роботі у режимі реального часу [3-10].

Вид нейромережевого алгоритму	Точність аналізу даних
Нейромережа прямого поширення	95%
Нейромережа на основі алгоритму рою частинок	96%
Нейромережа на основі ітераційного алгоритму рою частинок	97%
Нейромережа на основі еволюційного алгоритму рою частинок	98%

Слід зазначити, що проведення подальшого аналізу ефективності НММ вимагає співставлення ефективності нейромережевих алгоритмів у різних режимах роботи та навчання.

Висновки. В результаті аналізу було показано, що в основу комплексної моделі нейромережевого аналізу великих обсягів даних може бути покладена нейромережа прямого поширення типу багатосарового перцептронну або нейромережева архітектура на її основі. Визначено процедуру прогнозування та ідентифікації системних проблем, що виникають при обробці масивів даних у режимі реального часу. Зазначено, що для співставлення та аналізу результатів роботи нейромереж необхідно ввести такі базові критерії оцінки, як результативність, процент невдач, середньо квадратична похибка та коефіцієнт змішаної кореляції.

1. Buscema, Massimo, 2013. Supervised Artificial Neural Networks: Backpropagation Neural Networks. *Intelligent Data Mining in Law Enforcement Analytics: New Neural Networks Applied to Real Problems*. pp. 119–135. doi:10.1007/978-94-007-4914-6_7.
2. Buscema, Massimo, 2015. Supervised Artificial Neural Networks: Backpropagation Neural Networks. *Intelligent Data Mining in Law Enforcement Analytics*. Nov. 2012, Bangalore, P.; Tjernberg, L.B. An artificial neural network approach for early fault detection of gearbox bearings. *IEEE Trans. Smart Grid*, 6, pp. 980–987.
3. Illias, H. A., Chai, X. R., & Mokhlis, H., 2015. Transformer incipient fault prediction using combined artificial neural network and various particle swarm optimisation techniques. *Plos One*, 10(6), e0129363.
4. Zakaria, F.; Johari, D.; Musirin, I., 2014. Optimized Artificial Neural Network for the detection of incipient faults in power transformer. *In Proceedings of the 2014 IEEE 8th International Power Engineering and Optimization Conference (PEOCO)*, Langkawi, Malaysia, 24–25 March 2014; pp. 635–640.

5. Ahmed, M.R.; Geliel, M.A.; Khalil, A., 2013. Power transformer fault diagnosis using fuzzy logic technique based on dissolved gas analysis. In *Proceedings of the 2013 IEEE 21st Mediterranean Conference on Control & Automation (MED)*, Chania, Greece, 25–28 June 2013; pp. 584–589.
6. Setiawan, N.A.; Adhiarga, Z., 2012. Power transformer incipient faults diagnosis using Dissolved Gas Analysis and Rough Set. In *Proceedings of the 2012 IEEE International Conference on Condition Monitoring and Diagnosis (CMD)*, Bali, Indonesia, 23–27 September 2012, pp. 950–953.
7. Yang, H.T.; Huang, Y.C., 2016. Self-Organizing Networks. *Advanced Wireless Networks*, pp. 478–485., doi:10.1002/9781119096863.ch13.
8. Antoni, J., 2006. The spectral kurtosis: A useful tool for characterising non-stationary signals. *Mech. Syst. Signal Process.*, 20, pp. 282–307.
9. Borghesani, P., Pennacchi, P., Randall, R., Sawalhi, N., Ricci, R., 2013. Application of cepstrum pre-whitening for the diagnosis of bearing faults under variable speed conditions. *Mech. Syst. Signal Process.*, 36, pp. 370–384.
10. Wang, Z.; Liu, Y.; Griffin, P.J., 1998. A combined ANN and expert system tool for transformer fault diagnosis. *IEEE Trans. Power Deliv.*, 13, pp. 1224–1229.