Enhanced Multi-Label Classification Using Fuzzy Deep Neural Networks for Imbalanced Data

Dr. Harish B.R[1], Erat Krishnankutty Jishor[2], Mounika Akshaya[3]
[1]Lecturer, Engineering Department, College of Engineering & Technology, University of Technology and Applied Sciences, Shinas, Sultanate of Oman.
[2]Lecturer, University of Technology and Applied Sciences, Shinas, Sultanate of Oman.
[3]MS in Marketing, George Herbert Walker School of Business & Technology, Webster University, St Louis, Missouri, USA

Article Info

Article history:

Received Jan 9, 2025
Revised Feb 14, 2025
Accepted Mar 9, 2025

Keywords:

Deep neural networks (DNN)
Imbalanced multi-label datasets
Fuzzy inference systems
Data-driven inference
Multi-label classification

ABSTRACT

In classification issues where each data instance is simultaneously assigned many labels, a classification with multiple labels is a useful method for controlling uncertainty. These kinds of circumstances are common in real-world settings where judgments are based on ambiguous or crowded data and flexible categorization techniques are favored. Nonetheless, the issue of class imbalance is a feature shared by several multi-label datasets, where samples and the labels that belong to them are not distributed uniformly throughout the data space. To address the issue of class imbalance, we present a fuzzy logic-based multi-label classification method in this study. One or more labels could be used to convey this categorization challenge. Deep neural networks, which have shown themselves to be incredibly effective in such situations, are used to do away with the necessity for an expert to ascertain the logical principles of inference. The benefits and drawbacks of each strategy can be balanced by integrating deep neural networks with fuzzy inference systems. Accuracy and model application flexibility will be improved by the suggested system, especially when it comes to time limits brought on by causality of reported series of times. The suggested model performs better for continuous data classification than four baseline models, according to tests conducted on a classification with multiple labels dataset about present-day and voltage profiles of various home appliances.

Corresponding Author:

Dr. Harish B.R,
Lecturer, Engineering Department, College of Engineering & Technology, University of Technology and Applied Sciences, Shinas, Sultanate of Oman.
Email: harish.rudraiah@utas.edu.om

1. INTRODUCTION

Every observation in conventional unassisted machine learning techniques [1] is a member of one decision class since instances are often linked to a single label. Each instance of multi-label data can simultaneously belong to multiple classes, which is its distinguishing feature. In many application sectors, classification is an essential activity. The feature space of a classification issue, which is the compilation of all possible values for a certain collection of attributes, should ideally be well-separated to simplify the process of classifying a sample. The classic single-label categorization is one of the biggest and most widely used ML algorithms where this conceptual framework is implemented. This procedure has been effectively used in many different fields and is renowned for its adaptability [2]. Binary and multi-class classifications are two subclasses of the particular-label classification approach that involve acquiring knowledge given a collection of data points linked to a single label. Nonetheless, patterns from other classes may display comparable features in a variety of real-world situations, leading to overlapping regions in the area of features. This feature, which is the primary obstacle in multi-label classification, is frequently referred to as the class duplication problem [3].

Because multi-label problems are frequently high-dimensional, feature extraction makes developing and utilizing multi-label classification methods less complicated. Learning high-level characteristics from data that is accessible is one of the challenges of deep learning. Information granules that facilitate quicker learning are indicated by these high-level qualities. Since multi-label classification problems frequently require a large number of labels, obtaining high-level features is insufficient. Numerous writers have put forth multi-label categorization schemes that draw inspiration from deep learning methodologies [4]. A few of these systems depend on autoencoders, which enable unsupervised feature learning. This method effectively extracts the fundamental features from any supplied data, producing a smaller dataset that is well-coded.

Unbalanced data is a prevalent and difficult issue that impacts a classification model's learning process in any classification challenge [5]. Specifically, imbalance training is a common and intrinsic feature of many MLDs that influences how many classification algorithms train. There are three ways to look at the imbalance issue in an MLD: imbalance among the label sets, imbalance between labels, and imbalance within labels. When there is an imbalance within labels, there are typically a lot of negative samples and very few positive examples on each label. Each occurrence of an MLD is linked to several outlets or labels, thus it's typical for some of them to be majority labels and others to be minority labels—that is, some labels have a significantly higher number of favorable samples than others. The scarce frequency of label sets is the third kind of label imbalance that typically arises in MLD [6]. If a complete identify set is taken into consideration, the percentage of positive and bad samples for each group may be associated with the most common label arrays. Because of the label sparseness, MLDs usually have more frequent and separate label sets. This suggests that while some label sets might be regarded as majority cases, the rest label sets might be regarded as minority cases concurrently.

Three perspectives can be used to analyze the imbalance issue in a multi-label dataset: imbalance among label sets, imbalance between labels, and imbalance within labels [7]. The most common label sets may be linked to the proportion of positive versus negative assessments for each class when taking into account the entire label set. A traditional classifier could find it difficult to divide the relevant data into distinct classes or to discriminate between the classes when these difficulties arise [8]. This study addresses class disparity in a common real-world scenario involving the multi-label categorization of home appliances using a flexible deep neural network (FDNN). The suggested method is new in this regard in two ways. On the one hand, it addresses the issue of class imbalance by utilizing fuzzy logic and DNNs in tandem to automatically determine logical standards for inference without the requirement for human experience. Specifically, by dividing up historical data into sequences of symbols, the representational aggregate approximation (SAX) [9] technique enables dimensionality reduction. We take advantage of this potential. In this manner, training time can be decreased while maintaining precise outcomes.

[10] Real-world applications like energy communities and smart grids may benefit from this quicker retraining period, which also lowers overall investment costs and boosts societal penetration. GRUs efficiently handle time series classification by using data-driven methods to find and identify models that take into account the constantly changing development of the unknown material system generating the observed time series. The extended short-term memory network (LSTM), cynical CNN, one-dimensional flexible neural network with compression (FCNN), or the fuzzy variation of the CNN network, and a naive classifier (NC) employed as the starting point are the four models that are compared to the suggested technique [11].

2. LITERATURE REVIEW

Rosato et al. [12] proposed Multi-label classification with imbalanced classes by fuzzy deep neural networks. In classification issues where each data instance is simultaneously assigned many labels, multi-label classification is a useful method for controlling uncertainty. These kinds of circumstances are common in real-world settings where judgments are based on ambiguous or noisy data and flexible categorization techniques are favored. Nonetheless, the issue of class imbalance is a feature shared by several multi-label datasets, where samples and the labels that go with them are not distributed uniformly throughout the data space. To address the issue of class imbalance, we present a fuzzy logic-based multi-label classification method in this study. Deep neural networks, which have shown themselves to be incredibly effective in such situations, are used to do away with the necessity for an expert to ascertain the logical principles of inference. The benefits and drawbacks of each strategy can be balanced by integrating deep neural networks with fuzzy inference systems. The experimental findings demonstrate that selecting the ideal set of hyperparameters required to achieve high accuracy significantly affects calculation times, outweighing any potential advantages offered by SAX.

Zhuang et al. [13] introduced Multi-label learning based deep transfer neural network for facial attribute classification. The great efficacy of facial attribute categorization using DNN often requires a large volume of labeled data. To tackle the aforementioned problem, we propose FMTNet, a novel deep transfer neural network approach based on multi-label learning for facial assign classification, which consists of three sub-networks: Face detection Network (FNet), Multi-label learning Network (MNet), and Transfer learning Network (TNet). First, FNet is optimized for face detection by utilizing the Faster Region-based Convolutional Neural Network (Faster R-CNN) as its foundation. An efficient loss weight strategy is then created to specifically take use of the association between face attributes and their attribute grouping after MNet has been refined by FNet to anticipate multiple attributes with labeled data. Lastly, TNet is trained via unsupervised domain adjustment for unlabeled face attribute categorization, based on MNet. Tight coupling between the three sub-networks enables efficient classification of facial attributes. One novelty of the proposed FMTNet technique is that three of the systems of connections (FNet, MNet, and TNet) are constructed with a similar network structure. TNet in FMTNet, which is based on MNet, uses MK-MMD to predict facial features with unlabeled data. The suggested approach is broad and applicable to various computer vision problems since it blends transfer learning and multi-label learning.

Tarekegn et al. [14] presented a review of methods for imbalanced multi-label classification. An expansion of the conventional single-label classification, multi-label classification (MLC) assigns multiple labels to each data instance at the same time. MLC's broad range of application sectors has contributed to its increased significance in recent years. However, a class disparity, in which specimens and labels are distributed unevenly over the data space, has become an essential characteristic of many multi-label datasets. The asymmetry issue underlying MLC, which may be viewed from three perspectives—unbalance between label sets, imbalance within labels, and imbalance among labels—presents a barrier to multi-label data analytics. We provide a summary of the approaches to the imbalance problem in multi-label data by aggregating the literature. Each of these strategies has limitations, even though a number of them have shown promise in handling unbalanced lessons, names, and labeling sets in MLC. For example, the imbalance problem across label sets cannot be solved by methods intended for dealing with the imbalance problem between labels.

Shalaginov & Franke [15] suggested a deep neuro-fuzzy method for multi-label malware classification and fuzzy rules extraction. Contemporary malware poses a growing risk to organizational security as well as personal privacy. While most machine learning techniques perform well when it comes to the overall benign-malicious classification, they are unable to differentiate between numerous classifications. Deep learning is effective for Deep Neural Networks one example of such an application. To describe each malware family independently, it is impossible to extract an intelligible model. Therefore, this study has many objectives. We investigated multinomial sorting of malware employing static and dynamic behavioral metrics taken from the Microsoft Windows Portable Executables (PE32) infection dataset as of the last day of 2015. Through the use of deep learning, this technique aims to increase the accuracy of the traditional Neuro-Fuzzy methodology. Neural Networks with Deep Learning have recently seen costly use in a variety of applications. A verbal model explaining the choice being made, however, cannot be extracted. Therefore, we suggested the Deep Neuro-Fuzzy method, which outperforms DNN at the same degree of abstraction in terms of accuracy.

Bello et al. [16] proposed a deep neural network to extract high-level features and labels in multi-label classification problems. In deep neural networks, pooling layers assist cut down on redundancy and parameter count without requiring extra learning procedures. These operators are specifically designed to reduce feature space, even though they can handle both single-label and multi-label situations. In this work, we propose a deep neural architecture for multi-label classification issues that uses bidirectional association-based pooling layers for obtaining high-level features and labels. Our method finds unique pairings of neurotransmitters that will be combined to form pooled neurons using an association function. We propose to measure collaboration values in the primary pooling layer based on the correlation calculated by Pearson connecting each variable. Furthermore, we suggest an iterative method that eliminates the requirement to recalculate the correlation matrix in order to estimate the association degree of pooling neurons in deeper layers. This deep neural architecture's primary benefit is its ability to extract advanced characteristics and labels from datasets that lack a particular topological structure. The likelihood of creating networks with open inference models rises with the extraction of fundamental features and labels. Networks involving fewer parameters undoubtedly aid in achieving this goal, as these methods frequently have exponential algorithmic complexity.

3. PROPOSED SYSTEM

By aligning non-linearly separable classes with a separable hyperplane, each additional DNN layer introduces non-linearity. Deep feature mapping: When working with multinomial data, the primary problem with shallow single-layer neural architectures is their incapacity to distinguish samples from other classes due to shared characteristics.

We employed 15 MLC samples from the RUMDR repository to run the simulations [3]. Table 1 shows that the number of features in these problems ranges from 72 to 2,150, the frequency of instances from 207 to 368,638, and computing number of identifiers from 4 to 400. Additionally, we present the standard deviation of maximal absolute correlation that is calculated when features and labels are subjected to the association-based pooling approach. A standard classification problem is recognizing a specific finding from a preset set of categories. On the other hand, classification issues in real-world settings might be considerably more complex, involving data objects that concurrently belong to several classes. Setting precise borders between regions to categorize every observation is impractical in these situations.

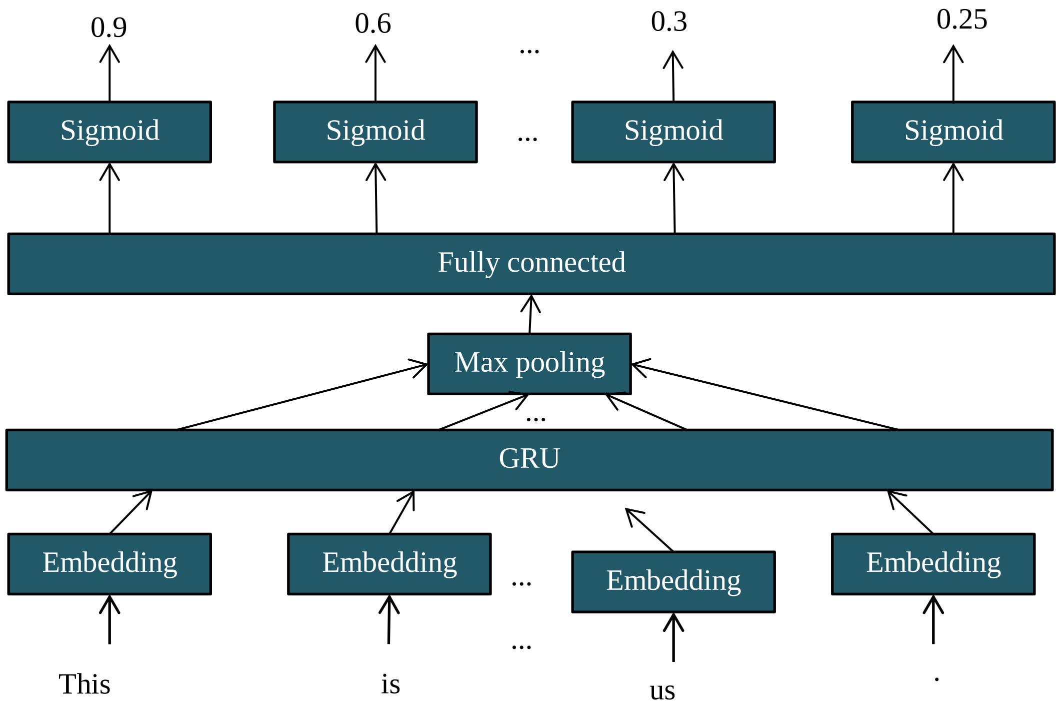

Figure 1. Proposed structure of Multi-label classification

Figure 1 illustrates the fundamental framework of multi-label categorization. In this part, we use several MLC problems to assess the suggested network architecture's performance. More precisely, we will investigate (i) the accuracy and quantity of top-level characteristics and labels produced by the model when the association thresholds are changed, (ii) the decrease in the problem accountability, and (iii) the accuracy loss brought on by using the features that were collected and labels. This classification problem can be solved by splitting the distance Γ into C discrete decision regions, which are as follows:

$$\Gamma = \Gamma_1 \cup \Gamma_2 \ldots \cup \Gamma_c \,,$$
$$l,, n = 0,1,2 \ldots c \qquad (1)$$

Where class n's region of decision making is denoted by Γ_n. On the other hand, real-world classification issues could be far more complex, encompassing informational objects that concurrently fall under several different classifications. Setting precise borders between regions to categorize every observation is impractical in these situations. Here, membership-based services $\mu_n(x)$, with $n = 0, 1, 2, 3$, transfer internal humidity range x to the phrases "COLD," "WARM," and "HOT." Despite this, it's crucial to remember that multilabel classification datasets typically have unbalanced classes, which makes the task more challenging. This work proposes an FDNN model that combines fuzzy logic and deep learning techniques to address this issue. In time series classification, the uncertain technique is crucial for addressing the intrinsic issue of class crossover, even if GRUs are successful when developing long-term relationships. This is particularly crucial since historical records that display the voltage and current values of multiple appliances may have profiles in common.

From the formal point of view, the neural network uses linear associative framework $g(S; \emptyset)$, depending on some parameter vector $\emptyset$ determined while training, in order of any intake time series $S_m(t)$, the variables $\gamma_n[S_m(e)], n = 1 \ldots\ldots C$, occur:

$$\begin{bmatrix} \mu_1 \\ \mu_2 \\ \cdot \\ \cdot \\ \mu_n \end{bmatrix} = g[S_m(e)] \qquad (2)$$

An overview of FDNN's operation is given by equation (2), which takes a duration series$S_m(e)$as its input and outputs the expected fuzzy values.

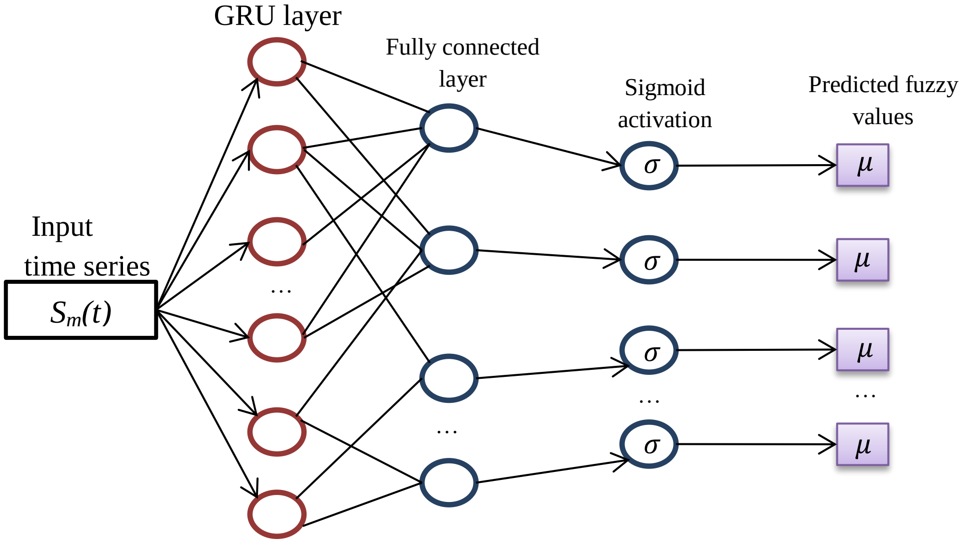

Figure 2. Proposed Multi-Label Classification Using Fuzzy Deep Neural Networks for Imbalanced Data Architecture

The suggested model, which is an updated version of the framework described in other models, is condensed in Figure 2. The architecture's layers are:

GRU layer: It emulates the real GRU layer, which has hidden components and recognizes relationships among the time increments of the supplied time series. The component in question introduces or eliminates state-related data at any given step. The "candidate hidden state" gate creates, the "update" gate produces, and the "reset" gate provides the matching column vector. These three gates are used by the layer to regulate periodic updates. Three gates are used by the layer to regulate these updates: The column vector r_t is obtained from the "reset" gate; the "update". The gate yields z_t, whereas the "candidate hidden state" produces $\hat{h}_t$. If st is an amount of $S_m(t)$ at the point in time t, the following equations demonstrate how a GRU cell functions (for simplicity's sake, the subscript m is removed):

$$\hat{h}_t = (1 - z_t) \odot \hat{h}_{t-1} + z_t \odot \hat{h}_t \qquad (3)$$

Fully connected layer: It is a typical forward feedback layer that connects the classes' C outcomes using the H replies from the preceding layer. The stratum adds a bias component after multiplying the input data by its proportion to the weight of the matrix.

Sigmoid layer: Every signal can be regarded as a vague affiliation level in the [0; 2] range since it gives each neuron in the fully connected layer a value that is sigmoid. This function is similar to:

$$f_{(y)} \frac{1}{1 + e^x} \qquad (4)$$

The following method can be used to discretize the length of series S once the breakpoints have been obtained. A time series is first subjected to a PAA. The sign an is applied to any PAA values that fall below the greatest modest breakpoint; the number b is applied to those that are smaller than the next smallest breakdown but larger than or equal to the lowest breakage; and so on.

The publicly accessible Plug-Level Appliance Identifying Dataset (PLAID) is used to evaluate the suggested methodology. One comprises high-frequency (40 kHz) voltage as well as current observations from 60 households in Pittsburgh, Pennsylvania, in the United States, using ten different types of appliances. Every appliance category in the dataset, which was collected in the summer of the year 2013, included multiple examples of various makes and models. A 1-second window that included the stable state of operation and, if applicable, the startup transition was taken from each of the initial three to 6 measurements that were taken for every appliance during the post-processing phase. CNNs have shown themselves to be effective tools in the deep machine-learning field for problems involving images. The authors presented an integrated structure for multilabel image recognition that models the label co-occurrence relationship in a hybrid image/label embedding space using CNNs and recurring neural networks. Since the refrigerator in the latter instance might typically be moved to another room due to space constraints or if there are multiple refrigerators in a single home, its degree of membership in this class is 0.3.

4. RESULT AND DISCUSSION

The practical considerations and results about the implementation of the suggested Multi-Label Classification Using Fuzzy Deep Neural Networks for Imbalanced Data architecture are presented below. Specifically, we employ: a standard CNN and its fuzzy variant, FCNN, with F filters; a neural network called LSTM with H hidden units, the industry standard for using time series in deep learning models; with a Naive Classifier (NC), an approach to classification that is employed as a baseline since it makes few assumptions about the problem at hand. The following three approaches are used to establish a reliable baseline: Make a random guess prediction, and Forecast a class within the training set that is chosen at random. Make a clear majority class prediction from the sample set. The ADM algorithm with a gradient decay factor of 0.9 and a rapid stopping criterion to limit the amount of repetitions is used to train all of the DNNs under consideration. In particular, the training process continues for 50 consecutive iterations until the loss function does not decrease. When there is little loss after the training process, the weights are returned to their initial values.

During the course of training, an inaccurate loss equation is employed, and the Root Mean Squared Error (RMSE) statistic collected over the samples is used to assess the final network performance:

$$RMSE = \frac{1}{W}\sum_{m=1}^{W}\sqrt{\frac{1}{E}\sum_{n=0}^{E}\left(t_{n,m} - \sigma_{n,m}\right)^2} \qquad (5)$$

In equation 5, where W is the total amount of samples in the one being tested set (i.e., W = 176), $t_{n,m}$ indicates the desired ambiguous adherence to the n-th class of the duration of the series $[S_m(t)]$, in accordance with the earlier illustration. $\sigma_{n,m} = [S_m(e)]$ is the correct value that the selected DNN classifier estimates.

The total quantity of layers was selected based on earlier models. While Deep NF utilized 100 trials for deep learning and 10 epochs for adjusting fuzzy rules, DNN employed 500 training epochs. The learning rate for both approaches was 0.3. It is evident that the suggested Deep NF performs better than DNN at the same degree of abstraction, whereas the standard Simple NF performs noticeably worse. However, when it comes to weight values, our approach has a larger computational complexity. The same logic holds for CNN, even though the decline in categorization accuracy is essentially insignificant. They are fairly comparable when seen from the perspective of the standard variations, demonstrating strong stability across all models. When considering computing occasions, each model will have faster training times, with the exception of the network CNN and FDNN (but exclusively for the "self-sufficient" method). It will lead to the extraction of further regulations. We show the classification accuracy achieved by a defuzzification method to give a more thorough assessment of the model's performance. This entails giving each pattern, taking into account both $t_{n,m}$ and $\sigma_{n,m}$, exclusive assignment to the most significant imprecise participation value for the classification label. The mean performance for each appliance for the first and second groups, as calculated using the proposed FDNN, is listed in Table 1. Specifically, it is clear how the ideal outcomes found in Table 1 correspond to the anticipated fuzzy descriptors that are more like the actual ones.

Table 1. Average performance

Examples	**First set**		**Second set**	
	RMSE	Accuracy	RMSE	Accuracy
Fridge	0.345	90	0.145	100
Fan	0.643	76	0.139	74
Laptop	0.723	46	0.452	100
Heater	0.1356	91	0.359	45

Table 1 displays the worst findings for the poorly forecasted affiliation degrees related to the heater and refrigerator. On the other hand, Heater's expected fuzzy values differ significantly from the actual ones, especially when considering the "Vehicles assisted" and "Non-motor assisted" categories. The suggested model finds it difficult to accurately forecast the actual values in certain situations. NC-A, NC-B, and NC-C are three completely different methods that were applied to the foundation sequence (NC). Again, the "Data" option selects whether the analysis considers the original strands or their SAX conversion. Specifically, the latter assumes nearly the same value when taking into account CNN, FDNN, and FCNN, with a roughly 9% relative improvement over the LSTM, which is unable to offer the same performance.

Table 2. Average RMSE, Accuracy, and Training time

Insight	**Models**	**RMSE**	**Accuracy(%)**	**Training period(s)**
Initial	FDNN	0.245 ± 0.0213	100	187
	FCNN	0.269 ± 0.047	74	23
	CNN	0.852 ± 0.236	100	96
	NC-A	0.914 ± 0.039	45	23
	NC-B	0.942 ± 0.24235	100	765
SAX	FDNN	0.287 ± 0.24235	45	188
	FCNN	0.946 ± 0.24235	90	124
	NC-A	0.342 ± 0.24235	32	56
	NC-B	0.179 ± 0.24235	100	23

Table 2 demonstrates that the proposed FDNN yields the most beneficial results in terms of RMSE, whereas CNN is the best method in regards of classification accuracy when considering the original data. DNN performance is determined by the amount of non-stationary adaptation to the sample data, which is defined by the total number of hidden layers used in the approach. We examined the classification accuracy and DNN training time about the number of hidden layers to assess the suitability of deep learning algorithms for these kinds of issues.

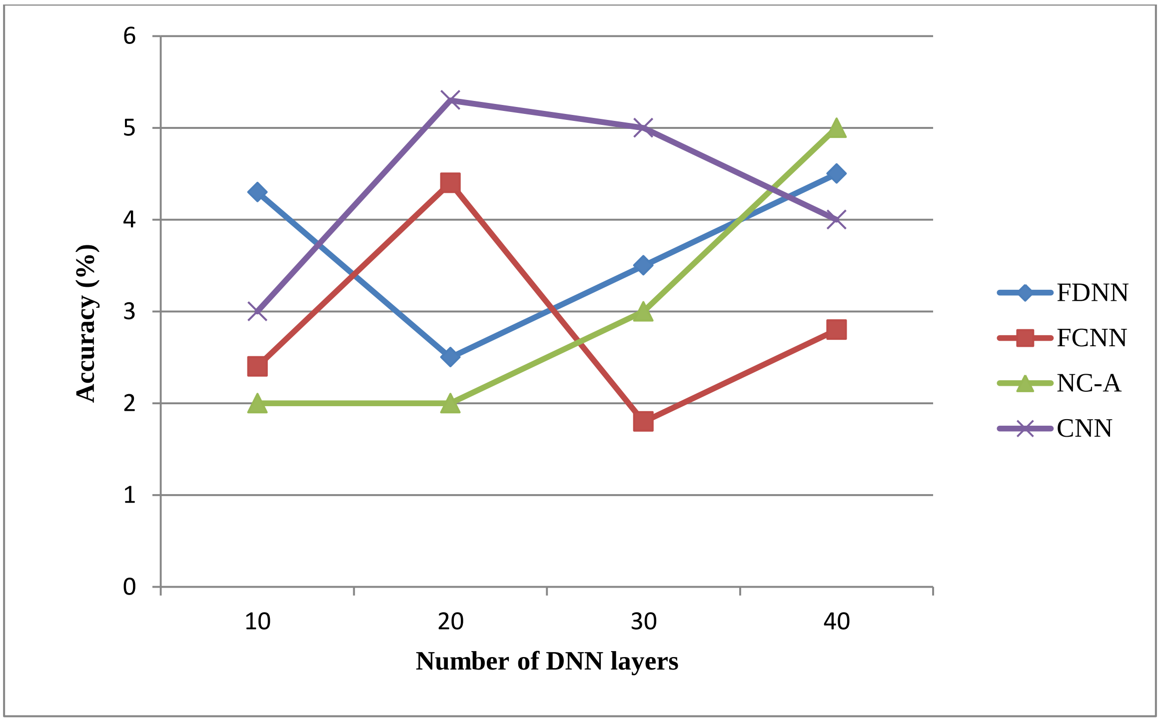

Figure 3. Comparison of Accuracy

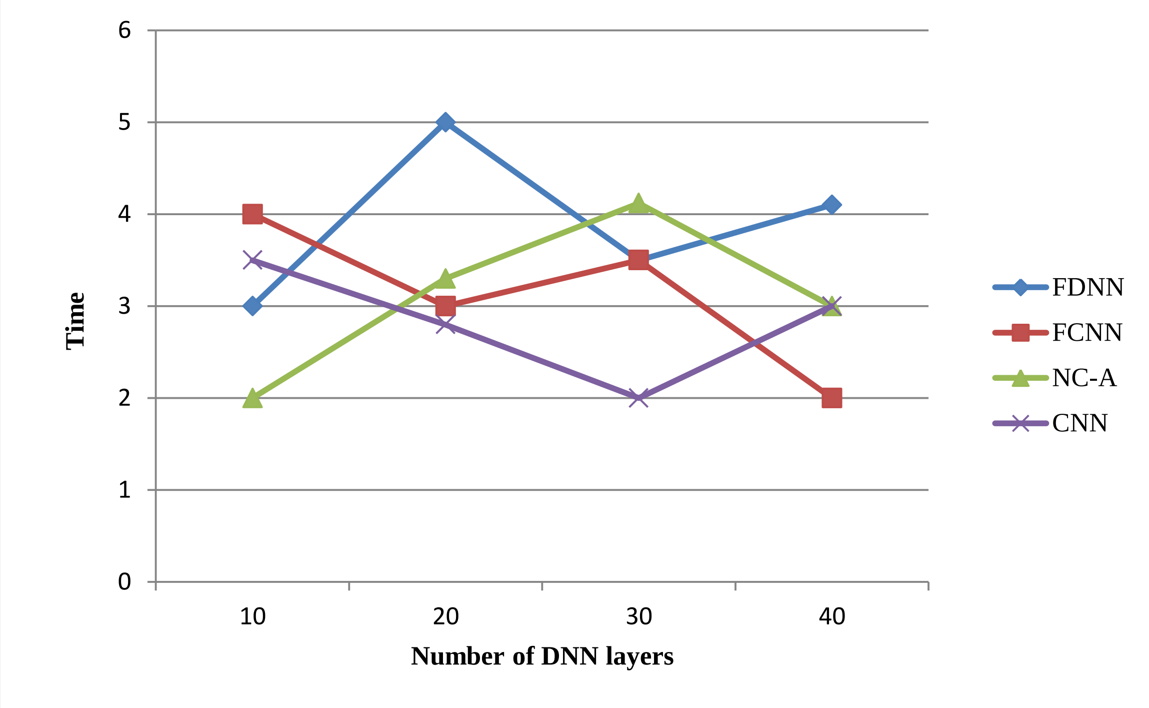

Figure 4. Comparison of Training time

Figure 4 displays a training time with layers up to 10,000, whereas Figure 3 shows how the quantity of layers (1–10) affects the DNN's accuracy. The suggested model is more accurate and requires less training time than previous models, and the RMSE values have been altered in line with this. As can be shown, attacker-developed evasion techniques exacerbate multinomial malware classification issues; hence, DNN cannot attain 100% accuracy but will do its best. Performance is improved by adding up to ten hidden layers; but, additional increases hardly justify using a perceptron with more than one layer. In actuality, our 20-layer DNN had the highest accuracy during the testing, properly classifying 92.043% of malware samples across 10 families.

5. CONCLUSION

A potential remedy for the multi-label classification strategy with unbalanced classes is offered in this work. The assignment pertains to the real-world issue of categorizing domestic appliances according to the time series linked to their profiles of voltage and current. Using an FDNN, which combines the concepts of neural networks alongside deep learning, solves the problem. The earlier method is used to address the overlapping issues of numerous classes, whilst the latter one is used to carry out the real multi-label categorization. The accuracy of the proposed approach is compared with four other models: CNN, FCN, LSTM, and an inadequate classifier used as a basis for comparison. In terms of reliability and robustness, the proposed FDNN performs better than any existing DNN. Faster training times without sacrificing accuracy are made possible by the time series' metaphorical representation. Beyond any potential benefits provided by SAX, the experimental results show that choosing the optimal set of hyperparameters needed to attain high accuracy has a considerable impact on calculation times. To better comprehend and (potentially) lower prediction errors, it is crucial to conduct a more thorough analysis of the time series itself.

REFERENCES

[1] Ouda, O., AbdelMaksoud, E., Abd El-Aziz, A. A., & Elmogy, M. (2022). Multiple ocular disease diagnosis using fundus images based on multi-label deep learning classification. *Electronics*, *11*(13), 1966.

[2] Elujide, I., Fashoto, S. G., Fashoto, B., Mbunge, E., Folorunso, S. O., & Olamijuwon, J. O. (2021). Application of deep and machine learning techniques for multi-label classification performance on psychotic disorder diseases. *Informatics in Medicine Unlocked*, *23*, 100545.

[3] Puri, V., Kataria, A., & Sharma, V. (2024). Artificial intelligence-powered decentralized framework for Internet of Things in Healthcare 4.0. *Transactions on Emerging Telecommunications Technologies*, *35*(4), e4245.

[4] El-Hajj, M. (2024, December). Cybersecurity and Privacy Challenges in Extended Reality: Threats, Solutions, and Risk Mitigation Strategies. In *Virtual Worlds* (Vol. 4, No. 1, p. 1). MDPI.

[5] Chen, L., Wang, Y., & Li, H. (2022). Enhancement of DNN-based multilabel classification by grouping labels based on data imbalance and label correlation. *Pattern Recognition*, *132*, 108964.

[6] Zhao, T., Zhang, Y., & Miao, D. (2022). Intuitionistic fuzzy-based three-way label enhancement for multi-label classification. *Mathematics*, *10*(11), 1847.

[7] Tian, D., Li, F., & Wei, Y. (2025). Interval type-2 fuzzy neural networks for multi-label classification. *Knowledge-Based Systems*, 113014.

[8] Bi, L., Feng, D. D., Fulham, M., & Kim, J. (2020). Multi-label classification of multi-modality skin lesion via hyper-connected convolutional neural network. *Pattern Recognition*, *107*, 107502.

[9] Li, Z., Xu, M., Yang, X., & Han, Y. (2022). Multi-label fundus image classification using attention mechanisms and feature fusion. *Micromachines*, *13*(6), 947.

[10] Banik, S., Lauri, M., & Frintrop, S. (2018). Multi-label object attribute classification using a convolutional neural network. *arXiv preprint arXiv:1811.04309*.

[11] Xie, S., Hou, C., Yu, H., Zhang, Z., Luo, X., & Zhu, N. (2022). Multi-label disaster text classification via supervised contrastive learning for social media data. *Computers and Electrical Engineering*, *104*, 108401.

[12] Succetti, F., Rosato, A., & Panella, M. (2025). Multi-label classification with imbalanced classes by fuzzy deep neural networks. *Integrated Computer-Aided Engineering*, *32*(1), 25-38.

[13] Zhuang, N., Yan, Y., Chen, S., Wang, H., & Shen, C. (2018). Multi-label learning based deep transfer neural network for facial attribute classification. *Pattern Recognition*, *80*, 225-240.

[14] Tarekegn, A. N., Giacobini, M., & Michalak, K. (2021). A review of methods for imbalanced multi-label classification. Pattern Recognition, 118, 107965.

[15] Shalaginov, A., & Franke, K. (2017, November). A deep neuro-fuzzy method for multi-label malware classification and fuzzy rules extraction. In *2017 IEEE symposium series on computational intelligence (SSCI)* (pp. 1-8). IEEE.

[16] Bello, M., Nápoles, G., Sánchez, R., Bello, R., & Vanhoof, K. (2020). Deep neural network to extract high-level features and labels in multi-label classification problems. *Neurocomputing*, *413*, 259-270.