International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.6593
***Article Received**: 22 February 2023 **Revised**: 20 March 2023 **Accepted**: 13 April 2023*

Financial Fraud Detection using Improved Artificial Humming Bird Algorithm with Modified Extreme Learning Machine

V. Rama Krishna[1], Sekharbabu Boddu[2]
[1]Phd Research Scholar ,Department of CSE,
KLEF,
Vaddeswaram, AP,India
mekrishnait1984@gmail.com
[2]Profressor ,Department of CSE,
KLEF,
Vaddeswaram, AP,India
sekharbabu@kluniversity.in

Abstract— More and more industries, including the financial sector, are moving their operations online as internet usage continues to rise at an exponential rate. As a result, financial fraud is on the rise in all its guises and in all parts of the world, causing enormous economic damage. The purpose of financial fraud detection systems is to identify potential dangers, such as unauthorised access or unusual attacks. In recent years, this problem has been attacked using a variety of machine learning and data mining methods. Aalgorithms, on the other hand, are better able to deal with only a small quantity of labelled data and a large amount of unlabeled data, making them useful in situations where it would be impractical to rely solely on supervised learning algorithms to train a good-performing classifier. In this research, we propose a Semi-supervised Extreme Learning Machine (SKELM) built on top of the weighted kernel, which we call SELMWK. For the purpose of detecting financial fraud, this research proposes an enhanced artificial hummingbird algorithm (IAHA). The algorithm combines two essential techniques to enhance its capacity for optimisation. To begin, the Chebyshev chaotic map is used to seed the first population of artificial hummingbirds, which boosts the population's overall ability to do global searches. Second, the guided foraging phase incorporates the Levy flight to enlarge the search field and forestall early convergence. The experimental results demonstration that the suggested technique recovers the Internet monetary fraud detections.

Keywords- Financial fraud; Improved artificial hummingbird algorithm; Semi-supervised Extreme Learning Machine; Chebyshev chaotic map; Weighted kernel.

I. INTRODUCTION

Before 1996, when Citibank and Wells Fargo Bank in the United States of America developed the first online banking application, clients could only transact business with their banks in person [1]. Credit card use online spread as a consequence of the widespread availability of online banking. E-commerce, online payment systems, telecommuting, online banking, and social networking are just some of the services that have proliferated and become increasingly popular in this area over the past decade [2]. Since then, fraudsters have ramped up their attacks on e-commerce transactions made through a variety of online payment systems [3]. Advances in digital technologies, especially for currency transactions, have altered the way in which people handle their finances on a daily basis. There has been a dramatic shift from traditional cash registers to online payment processing in recent years [4]. The use of technology in digital transactions has proved revolutionary for many economies, allowing them to maintain productivity and competitive advantage [5]. Customers can conduct banking and financial transactions, including the use of credit cards, from the convenience of their own homes or offices thanks to internet banking and other online transactions.

For a long time, fraud detectors and fraudulent transactions have worked hand in hand. In the modern era, especially the Internet, fraudulent transactions are rampant and account for the vast majority of monetary losses [6]. In 2019, the economy lost over $28 billion due to transaction fraud, while in 2020 and 2021, the losses increased to $30 billion and $32 billion, respectively. With an estimated $34 billion lost to fraud worldwide in 2022, financial and banking service providers may benefit from an automatic fraud deterrence solution [7]. Con artists that engage in fraudulent transactions do thus to enrich themselves monetarily or otherwise. They prey on the defenseless using methods including identity theft, phishing, and financial system manipulation. Despite the prevalence of electronic money in industrialized nations, the cost of payment

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.6593
***Article Received**: 22 February 2023 **Revised**: 20 March 2023 **Accepted**: 13 April 2023*

fraud is on the rise [8-10]. The most typical forms of fraud are card-not-present and account takeover. Individuals can reduce their risk of being a victim of fraud by being cautious with their personal information, maintaining up-to-date security measures, and remaining wary of suspicious requests. Administrations and law implementation are cracking down on fake activity, while financial institutions are attempting to increase security and educate customers.

Financial institutions that deal with credit cards or internet transactions must employ automated fraud detection systems. This not only builds confidence among buyers but also decreases losses. Advanced machine learning models [13] may now be used to detect fraud, thanks to the proliferation of both large data [11] and AI [12]. In fact, state-of-the-art data mining and methods are the backbones of today's very successful fraud detection systems. A binary classification model capable of differentiating fraudulent transactions from regular ones can be constructed from a data set that contains labelled transactions. The constructed model is applied to determining the legitimacy of incoming transactions. Using classification algorithms to identify potentially fraudulent financial transactions presents a number of difficulties. class imbalance (the ratio of fraudulent transactions to relatively tiny). Third, the interconnectedness of transactions over time; fourth, the occurrence of concept drift; that is, the fact that Class conditional distributions shift over time, necessitating regular updates to the classifier; fifth, the high dimensionality of the search space prior to feature pre-processing.

This research examines the algorithmic underpinnings of a semi-supervised classification problem in the field of fraud detection, wherein the data comes from the same domain. In this research, we present a weighted kernel-based technique for semi-supervised extreme learning machines (SELMWK). The SELMWK algorithm was designed with multi-core learning and semi-supervised extreme learning machines in mind. As a result, the semi-supervised extreme learning machines' poor operation stability caused by random hidden layers is improved, and the limited usefulness of single-core learning is avoided. Improved classification accuracy is another goal of this study, which is why the IAHA model is introduced as a new tool for feature selection. Here's how the rest of the paper is laid out: The history and related research of cutting-edge fraud detection techniques are presented in Section 2. The projected framework for fraud detection is originally labelled in Section 3. The experimental findings on the fraud dataset are obtainable in Section 4. The paper is finished with Section 5.

II. RELATED WORKS

Using gradient boosting and neural networks, Xu et al. [14] propose deep increasing decision trees (DBDT) as a new method for fraud detection. Construct a soft decision tree (SDT), a decision tree organized perfect with as nodes, and then use the concept of gradient boosting to create an ensemble of SDTs, combining the benefits of both traditional methods and deep learning. By doing so, we can enhance the representation learning capability of gradient boosting while keeping its interpretability intact by embedding neural networks into it. Also, during the model training stage, offer a compositional AUC maximisation technique to handle data imbalances at the algorithmic level, with the goal of increasing the rarity of identified fraud situations. Extensive experiments on a variety of real-world fraud discovery datasets demonstrate that DBDT can substantially improve performance while still retaining good interpretability. This https address hosts the code repository.

In order to identify potentially fraudulent financial transactions, Fanai and Abbasimehr [15] propose a two-stage system that makes use of supervised deep learning techniques and a deep Auto-encoder (DAE) as a representation learning approach. The proposed method has been shown to boost the efficiency of the used deep learning-based classifiers in experimental evaluations. In particular, when compared to their baseline performance on the original data, deep learning classifiers trained on the transformed -encoder significantly outperform. Additionally, deep Auto-encoder-generated models outperform both PCA-obtained dataset-generated and preexisting model-generated models.

The grouped trees and weighted ensemble procedure (GTWE) was developed by Xu et al. [16], who also established fraud prediction models for online loans using App behaviours based on logistic regression, extreme gradient memory (LSTM), and the GTWE procedure. The experimental findings demonstrate that the GTWE-based fraud prediction model achieves excellent classification effect and constancy with acceptable interpretability. Meanwhile, the fraud prediction model can detect clients with a fraud probability of up to 84.19%, suggesting that App habits have a significant role in predicting fraud in online loan applications.

New methods for detecting financial fraud using machine learning are proposed in [17] by Alwadain et al. A total of 6,362,620 synthetic transactions' transaction-level features were fed into a variety of machine-learning classifiers. We also examine the relationship between the various characteristics. In addition, a Conditional (CTGAN). The correctness of the projected predictor was shown to be higher than that of existing machine learning-based methods. Although though XGBoost had the highest accuracy score (0.999) out of all 27 classifiers, it only averaged a 0.998 accuracy when evaluated using fold cross-validation. The findings are significant for regulators and policymakers who want to create innovative and effective measures to mitigate the risk of financial fraud, and they are of particular relevance to financial institutions.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.6593
***Article Received**: 22 February 2023 **Revised**: 20 March 2023 **Accepted**: 13 April 2023*

Using representation learning, Van Belle et al. [18] offer CATCHM, based approach to credit card fraud detection (RL). We demonstrate the value of network RL for fraud detection by eliminating the need for manual feature engineering and taking into account the relational structure of transactions through clever network design, an effective inductive pooling operator, and careful shape of the classifiers in the downstream. CATCHM outperforms state-of-the-art methods in a comprehensive empirical evaluation on a real-world credit card dataset, demonstrating the approach's practical relevance for industry.

Inspired by generative adversarial networks (GANs), Teng et al. [19] propose a GAN-based framework they call BalanceGAN to detect online banking fraud on tremendously imbalanced data. To compensate for data discrepancies, we pre-train a fraud detection model on data from the generator before refining it with datasets. Experiments conducted on two real-world datasets demonstrate that BalanceGAN improves performance in Precision and Recall by more than 10% compared to traditional approaches for addressing imbalanced data.

Akinbowale et al. [20] used a simulated data explanatory research design to model the banking business. In order to detect the attendance of fraud and classify them into two groups hidden layer, and five hidden neuron layers was developed using a big data analytical approach called machine learning. Training, validation, and test datasets are generated from the input and output target samples automatically, and the confusion matrix is used to graphically display the proportions of correct and incorrect classifications. In addition, cluster analysis was applied to the fraud signs to classify them into related groups. These findings support the use of neural networks for risk assessment and fraud detection across three distinct degrees of internal fraud. The results of the confusion matrix corroborate this, showing that the majority of classes were correctly classified (95%) and that only 5% were misclassified. Additionally, the model indicates the viability of aggregating indicators of possible internal fraud. This research sheds light on the characteristics of internal fraud and offers a step-by-step plan for introducing an integrated forensic secretarial and big data knowledge outline for preventing such crimes. Forensic accountants should routinely add new data to machine learning models used for automatic classification and cluster analysis. Nothing is known about how to best use big data analytics and forensic accounting together to reduce the risk of fraud within financial institutions. It is hoped that the findings of this research would improve internal fraud prevention strategies and practises.

III. PROPOSED METHODOLOGY

In this part, we use a machine learning algorithm that selects features to detect financial fraud. Using the IAHA and SELMWK for classification helps increase accuracy. The following section will provide a brief summary of the dataset.

A. Dataset Description

The primary experimental dataset comes from a major Chinese online financial service provider. Following data cleaning and preparation, the dataset contains 192586 data samples, of which 4375 are fraudulent samples [21]. Almost sixty different pieces of information are contained in the dataset, including the following: the beginning level, payment records, financial state, status, etc. Not all the data fields are described in order to protect the privacy of certain information. The dataset is split into 8 subsets for the purpose of cross-validation, where the classification consequences of various machine learning models are compared. Each time there is an approximately 4:1 split between training and test data.

B. Data Pre-processing

Data validation, normalization, and division are three steps that must be taken before applying the model to the dataset.

1. Data validation

Data in the dataset is validated at this stage for errors like negative timestamps, blank values, and negative amounts.

2. Normalization

The perfect rescales the variables so that they fall inside the range [-1,1] to produce a more precise answer. To convert in the dataset to a standard scale without distorting the ranges of values or losing info, this process is essential. Equation 1 is used in the normalizing process.

$$x(i) = \frac{x(i)-\bar{x}}{s(x)} \qquad (1)$$

3. Dataset samples divide

The performance can be more accurately evaluated if the data is split into training and testing sets. Seventy percent of the data set is used for training in the proposed model, while the remaining thirty percent is rummage-sale for testing.

IV. FEATURE SELECTION USING IAHA

A. Brief Introduction of Artificial Hummingbird Algorithm (AHA)

The primary functions of AHA, which is a population-based metaheuristic procedure, are to model the foraging habits of guided foragers, territorial foragers, and migratory foragers, all of which are common among hummingbirds. Omnidirectional flight, diagonal flight, and axial flight are all modelled as they pertain to the foraging process. Simulating the hummingbird's exceptional memory capacity, an access table is built to instruct the algorithm towards global optimisation.

The three types of aviation expertise are as shadows:

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.6593
***Article Received**: 22 February 2023 **Revised**: 20 March 2023 **Accepted**: 13 April 2023*

Axial flight is defined as follows in order to bring the flight skill imitation into the d-D space::

$$D^{(i)} = \begin{cases} 1 & if\ i = randi([1,d])\ i = 1, \dots, d \\ 0 & else \end{cases} (1)$$

Diagonal flight is distinct as shadows:

$$D^{(i)} = \begin{cases} 1 & if\ i = p(j)\ \ P = randperm(k), k \in [2, [r_1(d-2)] + 1] \\ 0 & else \end{cases} (2)$$

Omnidirectional flight is distinct as shadows:

$$D^{(i)} = 1, i = 1, \dots, d\ (3)$$

For example, we can generate a random integer among 1 and d with randi([1,d]), a random permutation of integers among 1 and k with randperm(k), and a random sum in the range [0,1] with r1. In a d-dimensional space, a hyperrectangle contains a flight down the diagonal.

After a random set of solutions and a visit table have been generated, the AHA will begin its analysis. Guided or territorial foraging is carried out 50% of the time in each cycle. Guided foraging, which relies on nectar rates and a visit table, permits hummingbirds to travel in the direction of their preferred food sources. Hummingbirds can easily move to surrounding regions within their own range to forage, increasing the likelihood that they will discover suitable new food sources. After two iterations, there is a foraging migration. Everything is done interactively up until the stop rule is reached. The final result is an approximation of the global optimum, which is source with the fastest rate of nectar-refilling replacement.

Hummingbirds have their population size, n, randomly assigned to the number of food sources, n.:

$$x_i = Low + r \times (Up - Low) \qquad i = 1, \dots, n\ (4)$$

, where r is a chance vector in [0, 1], Low and Up are the bounds of the d-dimensional problem, and x i is the coordinate of the ith food source..

$$VT_{i,j} = \begin{cases} 0 & if\ i \neq j \\ null & i = j \end{cases} (5)$$

If the value of VT (i,j) is null, then the corresponding hummingbird is currently feeding at that source; if it is zero, then the ith hummingbird has recently visited the jth food basis.

Foraging with a guide: Hummingbirds have the aforementioned flight capabilities, allowing them to reach their goal food source and gather candidate food sources; thus, the following mathematical equation can be used to simulate guided foraging behaviour and candidate food sources.:

$$v_i(t+1) = x_{i,tar}(t) + a \times D \times (x_i(t) - x_{i,tar}(t))\ (6)$$

$$a \sim N(0,1)\ (7)$$

where x i (t) is the location of the ith food source for hummingbirds at time t, and x i,tar(t) is the location of the ith target food basis for hummingbirds at time t, and an is normally distributed with mean = 0 and deviation = 1.

A revised location for the ith food supply is as follows::

$$x_i(t+1) = \begin{cases} x_i(t) & f(x_i(t)) \leq f(v_i(t+1)) \\ v_i(t+1) & f(x_i(t)) > f(v_i(t+1)) \end{cases} (8)$$

where f () is the fitness value of the function. If the candidate refilling rate is higher than the present food source's, the hummingbird will continue to eat at the candidate food source resulting from Equation (6), as shown by Equation (8).

Hummingbirds may move on to new nectar sources after reaching a feeding destination. As a result, if a hummingbird runs out of its preferred food in one part of its area, it can simply relocate to another part of its territory. This is the mathematical equation used to model the territorial foraging strategies and potential food sources of hummingbirds.:

$$v_i(t+1) = x_i(t) + b \times D \times x_i(t)\ (9)$$

$$b \sim N(0,1)\ (10)$$

mean = 0, standard deviation = 1; b follows a normal distribution.

When food is in short supply in a hummingbird's typical territory, the bird will often travel great distances in search of suitable feeding grounds. A relocation coefficient is specified in the AHA algorithm. When the number of iterations surpasses a certain threshold, the hummingbirds in the food basis with the lowest filling rate will randomly relocate to a different food source anywhere in the search space. Eventually, the hummingbird will stop visiting the original source and begin feeding only on nectar from the randomly produced new source, at which point the hummingbird's migrating foraging behaviour can be described as follows.:

$$x_{wor}(t+1) = Low + r \times (Up - Low)\ (11)$$

where x_{wor} is the population's worst nectar replenishment food source.

B. Improved Artificial Map and Levy Flight (IAHA)

1. Chebyshev Chaotic Map

Chaos is a nonlinear system's natural progression towards an unpredictable state. Long-term behaviour with no defined timescale, caused by sensitive to beginning conditions.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.6593
Article Received: 22 February 2023 Revised: 20 March 2023 Accepted: 13 April 2023

Ergodicity is a hallmark of chaos, thus it stands to reason that searches conducted with chaotic variables would perform better than random ones [22].

The distribution range of the Chebyshev chaotic map is larger and more unchanging, and it can be dispersed throughout the interval [1, 1]. Any beginning value chosen for an iterated sequence with k 2 (where k is the order) will result in an uncorrelated, chaotic, and ergodic sequence up to this point. Here is a representation of the equation::

$$x_{n+1} = cos(karccosx_n) \qquad x_n \in [-1,1] \ (12)$$

The equation is used in this research to generate evenly dispersed points for initialising the position of artificial hummingbirds, which improves the initial population's global search ability and the algorithm's solution accuracy.

2. Levy Flight

Levy, a French mathematician, proposed the Levy distribution, a type of probability distribution, in the 1930s. Since then, numerous investigations of the Levy distribution have been conducted. It has been shown that many animals in the wild follow the Levy distribution in their foraging behaviour. Several random phenomena, including Brownian motion, random walks, and the Levy flight, adhere to the principle of the delivery. Levy flight is currently popular in the field of intelligent optimisation. The Cuckoo algorithm, for instance, uses Levy flight to determine where to travel next [23]. By including Levy flight into the AHA algorithm, the search space can be enlarged, making it simpler to avoid premature convergence.

Get the latest on Levy's flight status here::

$$x_i(t+1) = \begin{cases} x_i(t) + a \oplus Levy(\lambda) & f(x_i(t)) \le f(v_i(t+1)) \\ v_i(t+1) & f\big(x_i(t)\big) > f(v_i(t+1)) \end{cases} (13)$$

where x_i^t is the tth generation location of x_i, $\oplus$ is the dot increase, a is the step scope switch parameter, and $Levy(\lambda)$ is the chance search path, which contents:

$$Levy{\sim}u = t^{-\lambda}, 1 < \lambda \le 3 \ (14)$$

Its step size obeys the Levy delivery, and step scope s is intended as:

$$s = \frac{\mu}{|v|^{1/\beta}} \ (15)$$

where μ, v are normally distributed, defined as:

$$\mu{\sim}N(0, \sigma_\mu^2)$$

$$v{\sim}N(0, \sigma_v^2) \ (15)$$

Where,

$$\sigma_\mu = \frac{(1+\beta)(sin\frac{\pi\beta}{2})}{\frac{1+\beta}{2}\beta 2^{\frac{\beta-1}{2}}} \ (16)$$

$$\sigma_v = 1$$

to which b is typically set at 1.5.

Some financial variable star are discovered to be more significant than others for prediction, while others are found to have detrimental effects on categorization precision. Choosing the right set of financial indicators is crucial to the success of any learning algorithm, and a lack of care in this area often results in issues with false positives and false negatives. As a result, the right variables should be used to detect financial statement fraud. One can classify the primary financial inputs into 10 broad categories, including solvency, activity, profitability, EVA value. The resultant set of 58 financial pointers serves as the second-level variable quantity in this study.

Most of the determinants chosen agree with findings from other research. In addition, the ratio of net fixed assets to total assets is listed as one of the top fraud indicators with the logs of total debt, equity, and debt to equity. Profitability, structural ratios are also useful indicators for spotting fraud. Raw data must be cleaned and normalised before it can be utilised in a model since it may contain noise, distortion, or outliers that would otherwise be undetectable. If the value of an attribute is missing, we have eliminated the corresponding sample from the dataset.

In addition, when used as input to deep learning models, datasets with variables of different scales tend to produce subpar results. The features can be better used as input to algorithms when they are scaled and standardised such that they are all roughly the same size. It is also important to decrease data dimension to ensure the correctness and reliability of the data analysis and mining results due to the enormous reports.

C. Classification using Modified Extreme Learning Machine

The generalisation ability of the SKELM algorithm is dependent on the function, and the choice of the kernel function typically impacts the presentation of the whole SKELM model. The feature mapping of invisible layers is typically handled by a single kernel function in standard SKELM models. There are, however, downsides to this. Because the model's performance will vary depending on the data set, a single KELM is not particularly versatile.

This paper offers an enhanced semi-supervised technique called SELMWK that is based on the weighted kernel machine to increase the trained classifier's recognition effect, stability, and

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.6593
Article Received: 22 February 2023 Revised: 20 March 2023 Accepted: 13 April 2023

generalisation capacity. Benefiting from the semi- extreme learning machine's strengths while compensating for the single-kernel model's lack of generalisation capacity, this model is a win-win.

1. Improved Design of Kernel Function

Combining numerous kernel is a viable option for expanding the usefulness of a single kernel function. Several methods of mixing kernels shown that utilising more than one kernel yielded better results than using only one. This semi-supervised learning challenge makes use of the weighted kernel function computed using this calculation strategy after a large number of experiments. Moreover, the results are superior than those of single-kernel learning. Here's a quick rundown of four typical elementary kernel operations:

As stated in Eq. (17), a linear Kernel function takes the following form:

$$K(x_i, x_j) = x_i^T x_i \quad (17)$$

Hyperparameters c and d for a polynomial Kernel function are illustrated in Equation (18):

$$K(x_i, x_j) = (x_i^T x_j + c)^d \quad (18)$$

Equation (19) demonstrates the shape of the hyperparameter, the kernel of the radial basis function:

$$K(x_i, x_j) = exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (19)$$

The shape of the sigmoid kernel function is given by Equation (20), where a and c are hyperparameters.

$$K(x_i, x_j) = tanh(ax_i^T x_j + c) \quad (20)$$

A linear kernel function is a special case of the RBF kernel. It is a linear map. The RBF is less complicated than the poly kernel function because it needs fewer parameters to non-linearly transfer data to high-dimensional space. The complexity of the calculation is further exacerbated by the fact that when the value of the k endlessly small. In accordance with Mercer's theorem, a sigmoid cannot be the dominant structure. This study opts to use the RBF kernel as its primary kernel function, with weighted combination yielding the weighted kernel. Below are the measures taken in the calculation:

Next, build a weighted nonlinear charting function (x), where is a allowance limit and _1 (x) and _2 (x) are two separate nonlinear charting purposes, using the form indicated by Equation (21).

$$\Phi(x) = \theta\Phi_1(x) + (1-\theta)\Phi_2(x) \quad (21)$$

Then define a kernel function $K(x_i, x_j) = < \Phi(x_i), \Phi(x_j) >$, and bring Equ (10) into it to get Equation (22):

$$K(x_i, x_j) = \theta^2 K_1(x_i x_j) + (1-\theta)^2 K_2(x_i, x_j) + 2\theta(1-\theta)K_{1,2}(x_i x_j) \quad (22)$$

where $K_1(x_i, x_j) = < \Phi_1(x_i), \Phi_1(x_j) >, K_2(x_i, x_j) = < \Phi_2(x_i), \Phi_2(x_j) >$ is a common basic kernel function, $K_2(x_i, x_j)$ function.

When RBF is selected as the, their procedures can be uttered as Equations (23)–(25):

$$K_1(x_i x_j) = exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma_1^2}\right) \quad (23)$$

$$K_2(x_i x_j) = exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma_2^2}\right) \quad (24)$$

$$K_{1,2}(x_i x_j) = \left(\frac{2\sigma_1\sigma_2}{\sigma_1^2+\sigma_2^2}\right)^{\frac{N}{2}} exp\left(-\frac{\|2x_i - x_j\|^2}{\sigma_1^2+\sigma_2^2}\right) \quad (25)$$

where σ_1 and σ_2 are the limits of the RBF kernel purpose, and N is the measurement of the input data x.

The weighted kernel can be gotten by relieving Equations (23)–(25) into Equation (22), and kernel form is exposed in Equation (26):

$$K(x_i, x_j) = \theta^2 K_1(x_i x_j) + (1-\theta)^2 K_2(x_i x_j) + 2\theta(1-\theta)K_{1,2}(x_i x_j) \quad (26)$$

The basic weighted kernel function in this analysis is given by equation (26). It combines a standard kernel function with a cross kernel function to form its own unique kernel function. By eliminating the problematic of flattening variable info caused by the straight linear addition of single kernel functions, and by translating data to multiple feature spaces, it improves the classifier's presentation.

2. Semi-Supervised Weighted Algorithm Procedure

In an effort to jointly train a high-quality classifier using both labelled and unlabeled data, this paper proposes the SELMWK algorithm, which employs the manifold regularisation technique.

To obtain the hidden layers, SSELM employs explicit random mapping. This approach can be applied to a wide variety of glitches, but it suffers from low operational stability and necessitates intensive computational effort when dimensions. Both of the other approaches rely on kernel mapping, an implicit mapping technique that does not provide a straightforward way to determine the buried layer. In comparison to multi-core SKELM, single-core SKELM

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.6593
Article Received: 22 February 2023 Revised: 20 March 2023 Accepted: 13 April 2023

provides more reliable operation but fewer potential uses. The kernel allows SELMWK to examine the data from many perspectives and extract the most important information with reasonable application and operation stability. To test the label data, the trained classifier is created after the hidden layer is obtained by two constraints and the manifold regularisation loss purpose is minimised while the empirical is minimised. Below, we shall describe the SELMWK algorithm's calculating flow in greater depth. To begin, let's write out Equation (27) as the algorithm's optimisation problem statement.

$$\min_{f\varepsilon\mathcal{F}} \lambda L_m\big(f(X)\big) + CL\big(Y_L, f(X_L)\big) + \|f\|_H^2 \quad (27)$$

X=[X LT,X UT]T is the data, X L is the labelled data, X U is the un-labelled data, Y L is the true label of the labelled data, and f() is the association map, where and C are the consequence parameters. The X distribution of the data represents a geometric structure, and the first term is the loss function of various regularisation, which attempts to bring the projected outcomes into better agreement with this structure. The second part is an attempt to make the projected results more in line with the actual ones; this is the empirical loss function. In order to avoid the model becoming oversimplified, the third term acts as a regularizer.

Since the kernel mapping is an implicit nonlinear transformation of the input data, the value of X is most easily represented as Equation (28):

$$H = \Phi(X) \quad (28)$$

In the formula, $H = [H_L^T, H_U^T]^T$, $\Phi(\cdot)$ the matrix can be written as Equation (29) for a weighted nonlinear (x j), where. denotes the inner creation of vectors.

$$K = HH^T \quad (29)$$

Since the predicted labels $\tilde{Y} = H\beta, \tilde{Y} = [\tilde{Y}_L^T, \tilde{Y}_U^T]^T$, β is the network's weight matrix at output, the loss purpose of the first period of various regularisation in Equation (16), expressed as Equation (30):

$$L_m = Tr(\beta^T H^T LH\beta) \quad (30)$$

It is possible to express the second loss purpose in Equation (27) as Equation (31):

$$L\big(\Upsilon_L, f(X_L)\big) = \sum_{i=1}^{N_L} \|\tilde{y}_{Li} - y_{Li}\|^2 \quad (31)$$

where N L is the quantity of labelled information. The loss purpose of the additional element in Equation (27) can be obtained by simplifying Equation (31) to the form Equation (32):

$$L\big(\Upsilon_L, f(X_L)\big) = \|H_L\beta - \Upsilon_L\|^2 \quad (32)$$

In order to acquire the function that needs to be optimised, we plug Equation (32) and Equation (30) into Equation (27) to get Equation (33).

$$\min_{\beta} \frac{\lambda}{2} Tr(\beta^T H^T LH\beta) + \frac{C_0}{2}\|H_L\beta - \Upsilon_L\|^2 + \frac{1}{2}\|\beta\|^2 \quad (33)$$

where and C 0 are tuning parameters used to adjust the relative importance of these two variables. C 0 can be written as a diagonal matrix, like in Equation (34), to simplify further computations.

$$\Lambda_C = dig(C_0, C_0, \cdots, C_0, 0, \cdots, 0) \quad (34)$$

The number C0 in Eq. (34) represents the total sum of labelled samples, but the number 0 does not represent the total sum of unlabeled examples. Thus, Eq. (33) becomes Eq. (35):

$$\min_{\beta} \frac{\lambda}{2} Tr(\beta^T H^T LH\beta) + \frac{1}{2} Tr[(\Upsilon^* - H\beta)^T \Lambda_c (Y^* - H\beta)] + \frac{1}{2}\|\beta\|^2 \quad (35)$$

In the formula, $Y^* \in R^N$, where Rows 1 through NL include labelled data, but rows 2 through NL are false labels with no data. Equation (36) is produced by calculating the gradient of the solution to the optimisation problem posed by Equation (35).:

$$\nabla = \beta + H^T \Lambda_C (H\beta - \Upsilon^*) + \lambda . H^T LH\beta \quad (36)$$

Let∇ = 0, the production network is gotten as Equation (37):

$$\beta = H^T (\Lambda_C HH^T + \lambda LHH^T + I_N)^{-1} \Lambda c \Upsilon^* \quad (37)$$

When Equation (29) is substituted into Equation (37), the network's output weight matrix takes the form of Equation (38).

$$\beta = H^T (\Lambda_C K + \lambda LK + I_N)^{-1} \Lambda c \Upsilon^* \quad (38)$$

As a mapping, its precise form is unknown. Without knowing HT, there is no way to tell. It may not always be necessary to derive the network's output. The end yield of the SELMWK procedure is represented by equation (39)..

$$Y_{out} = [K(x_t, x_1), \ldots, K(x_t, x_N)](\Lambda_C K + \lambda LK + I_N)^{-1} \Lambda c \Upsilon^* \quad (39)$$

D. Post-applying the model

There are two parts to this process, and they are output acquisition and output evaluation. The results for the dataset will be displayed, and the existence of fraud will be determined. After then, the data will be analysed and assessed. Adjusting model parameters and conducting experiments on varying numbers of layers and iterations form the basis of the evaluation process. The objective is to determine which values of the adaptive parameters optimise prediction and processing speed.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.6593
***Article Received**: 22 February 2023 **Revised**: 20 March 2023 **Accepted**: 13 April 2023*

You should also experiment with other values for the number of layers and the sum of iterations to see what works best.

This research makes use of a dataset that includes a column of numbers called features. Each of these characteristics holds unique information, such as:

First, "time" indicates how long it has been between two separate transactions for the same account number.

Second, "Amount" means the condition of the quantities involved in the deal.

Thirdly, "V1 through V28" contain some delicate details.

The fraudulent scheme is referred to as "Class" in Having a 1 in a column indicates fraud, while a 0 indicates that none has occurred..

V. RESULTS AND DISCUSSION

Experiments are run in groups on a cluster of 30 similar computers, with one machine acting as the master and the others as the workers. There are 8 physical processor cores and 64 GB of RAM in each system. 7 with the latest versions of Java SE Expansion Kit and Scala 2.12 is the OS in use..

A. Performance Metrics

We measured the efficacy of our models in terms of both raw performance and how quickly they could be put into action. This allowed us to evaluate the efficacy and speed of our structures in relation to other state-of-the-art methods. Some of these metrics included area under the curve (AUC), F1, and kappa. Additionally, the k- validation test was used to determine performance. Our performance measurements are computed using Equations (40-44) while the focus of the execution speed measures was on training and prediction time.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (40)$$

$$Precision = \frac{TP}{TP+FP} \quad (41)$$

$$Recall(True\ positive\ Rate) = \frac{TP}{TP+FN} \quad (42)$$

$$F1Score = \frac{2TP}{2TP+FP+FN} = \frac{2\times Precision\times Recall}{Precision+Recall} \quad (43)$$

$$Cohen\ Kappa\ Score = \frac{Accuracy-P_{random}}{1-P_{random}} \quad (44)$$

The existing models are considered from [14-16] and ELM that uses various dataset for fraud detection. Therefore, the existing models are implemented with our dataset and consequences are averaged in Table 1 and 2.

Table 1: Validation Analysis of Proposed Model without IAHA

Model	Accuracy	Precision	Recall	F1-score	Cohen Kappa score
DBDT	85.29	0.861	0.853	0.850	0.6912
DAE	85.29	0.856	0.882	0.883	0.6996
GTWE	86.76	0.873	0.868	0.865	0.7233
ELM	88.23	0.885	0.882	0.882	0.7597
Proposed	91.17	0.912	0.912	0.912	0.7809

In the analysis of accuracy, the proposed model achieved 91.17%, where the existing models such as DBDT, DAE, GTWE and ELM achieved nearly 85% to 88%. The reason for poor performance is that all features are directly given to the classifiers for fraud detection. When the models are tested with precision, recall and F1-score, the DBDT achieved 86% to 85%, DAE achieved 85% to 88%, GTWE achieved 86% to 87% and ELM achieved 88%. But the proposed model achieved 91% of precision, recall and F1-score. When likening with all techniques, DBDT and DAE achieved 69% of Cohen kappa score, GTWE achieved 72%, ELM has 75% and proposed model achieved 78% of kappa score.

Table 2: Analysis of Projected Model with IAHA

Model	Accuracy	Precision	Recall	F1-score	Cohen Kappa score
DBDT	91.17	0.912	0.912	0.912	0.7809
DAE	92.64	0.928	0.928	0.928	0.8203
GTWE	92.64	0.930	0.926	0.926	0.8529
ELM	94.11	0.943	0.941	0.941	0.8825
Proposed	95.55	0.956	0.956	0.956	0.9118

When all models are tested with IAHA, each technique achieved better presentation in terms of precision, F1-score and kappa score. The influence of IAHA is high, even for DBDT, DAE and GTWE, for instance, the models achieved nearly 91% to 92% of accuracy, and F1-score and 80% to 85% of kappa score. ELM achieved 94% of recall and F1-score and 88% of kappa score. But the proposed model achieved 95% of accuracy, 91% of score, 95.6% of precision, recall and F1-score. Figure 1 to 5 presents the graphical analysis of projected model with existing techniques.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.6593
***Article Received**: 22 February 2023 **Revised**: 20 March 2023 **Accepted**: 13 April 2023*

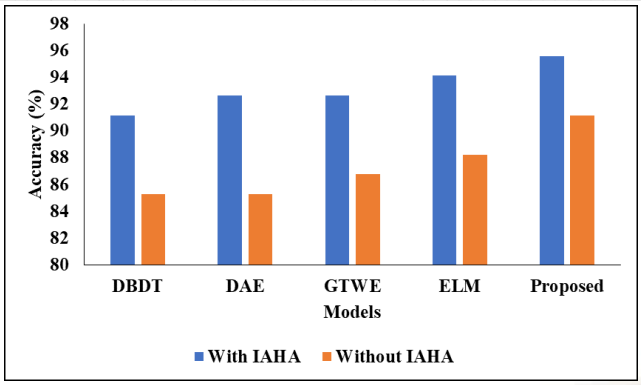

Figure 1: Accuracy Comparison

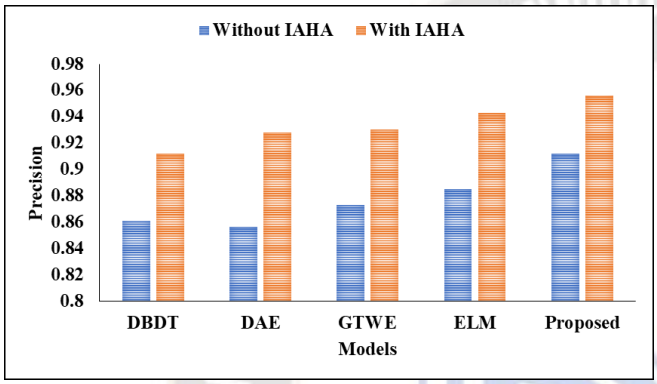

Figure 2: Analysis of projected classical in terms of Precision

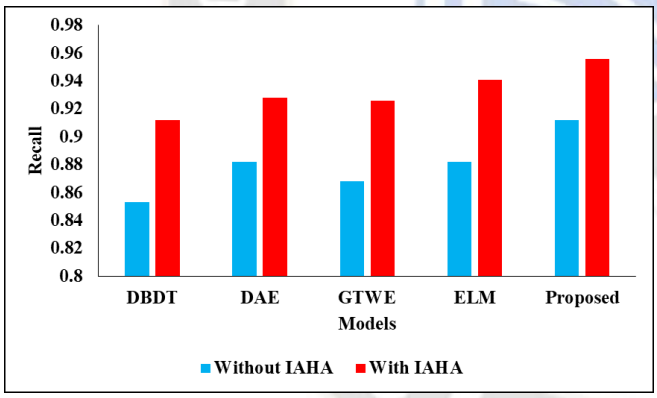

Figure 3: Recall Analysis

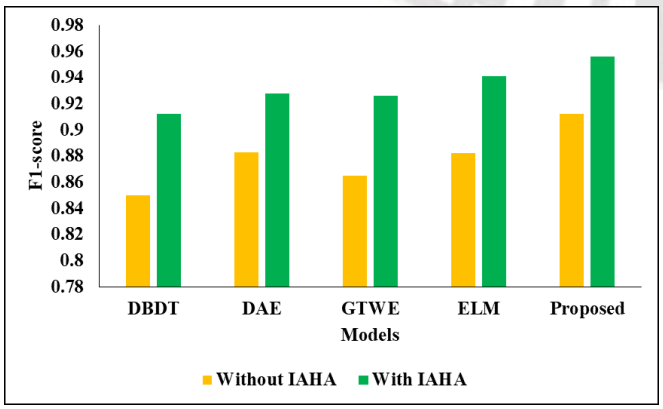

Figure 4: Graphical analysis of projected perfect in terms of F1-score

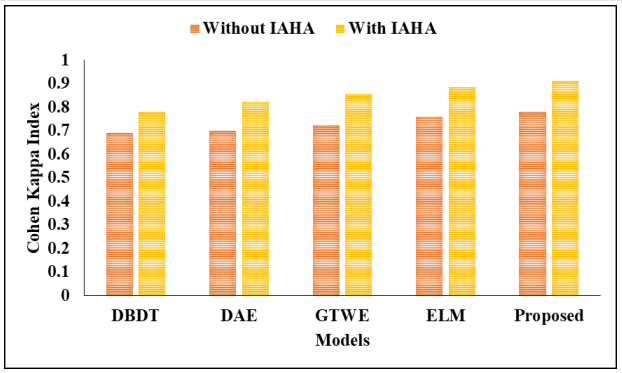

Figure 5: Cohen Kappa Score Analysis

VI. Conclusion

The effects of financial fraud on the financial system and its constituents are widespread. In recent years, the problem has been exacerbated by the widespread adoption of new technologies. In the era of big data, conventional methods no longer work. Commercial banks and other financial organisations have suffered massive losses as a result of Internet financial fraud instances. This article proposes a clever ML method for improving the effectiveness of financial fraud detections. Because of this, a model for the detection of financial fraud based on feature selection with enhanced ELM was built in the course of the work. In this work, we propose an enhanced AHA (IAHA) for use in selecting features. Enhanced alternatives to the traditional AHA include two new methods. Then, we used the Chebyshev chaotic map's population initialization to increase both the scope of our search and the precision with which we could find what we were looking for. Second, we improved the algorithm's convergence and stability by including Levy flight to steer foraging. Inspired by the two algorithms SSELM and SKELM, we explore manifold regularisation for classification and offer a technique, SELMWK, based on dangerous learning. Experiments measuring things like precision rate, accuracy, and kappa score demonstrate that the suggested method outperforms its competitors when it comes to learning and representing the properties of samples. It is proposed for future work to use deep learning with feature selection as a means of enhancing fraud detection classification accuracy..

References

[1] Srokosz, M., Bobyk, A., Ksiezopolski, B. and Wydra, M., 2023. Machine-Learning-Based Scoring System for Antifraud CISIRTs in Banking Environment. Electronics, 12(1), p.251.

[2] Alwadain, A., Ali, R.F. and Muneer, A., 2023. Estimating Financial Fraud through Transaction-Level Features and Machine Learning. Mathematics, 11(5), p.1184.

[3] Fanai, H. and Abbasimehr, H., 2023. A novel combined approach based on deep Autoencoder and deep classifiers for

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.6593
***Article Received**: 22 February 2023 **Revised**: 20 March 2023 **Accepted**: 13 April 2023*

credit card fraud detection. Expert Systems with Applications, p.119562.

[4] Mayo, K., Fozdar, S. and Wellman, M.P., 2023. Flagging Payments for Fraud Detection: A Strategic Agent-Based Model.

[5] GRADXS, G.P.B. and RAO, N., 2023. Behaviour Based Credit Card Fraud Detection Design And Analysis By Using Deep Stacked Autoencoder Based Harris Grey Wolf (Hgw) Method. Journal of Information Systems, 35(1), pp.1-8.

[6] Lukman, R.P. and Chariri, A., 2023. THE ROLE OF INTERNAL AUDITORS IN FRAUD PREVENTION AND DETECTION: EMPIRICAL FINDINGS FROM GENERAL BANKING. Diponegoro Journal of Accounting, 12(1).

[7] Strelcenia, E. and Prakoonwit, S., 2023. Improving Classification Performance in Credit Card Fraud Detection by Using New Data Augmentation. AI, 4(1), pp.172-198.

[8] Dammavalam, S.R. and Mukheed, M., Credit Card Fraud Detection Using Machine Learning.

[9] Dolu, U., 2023. A novel sampling technique and gradient boosting tree-based approach for cross-channel fraud detection.

[10] Sharma, K., 2023. Fraud Detection Model Using Semi-supervised Learning. In Soft Computing for Problem Solving: Proceedings of the SocProS 2022 (pp. 395-406). Singapore: Springer Nature Singapore.

[11] Okaka, J.M., 2023. Examination of fraud in the Ugandan banking sector and its prevention: a case study of Post Bank Uganda Limited (Doctoral dissertation, Makerere University).

[12] Zakaria, P., 2023. Financial Inclusion to Digital Finance Risks: A Commentary on Financial Crimes, Money Laundering, and Fraud. In Financial Technologies and DeFi: A Revisit to the Digital Finance Revolution (pp. 123-130). Cham: Springer International Publishing.

[13] Owiti, S.O., Ogara, S. and Rodrigues, A., 2023. CONTRIBUTING FACTORS TO MOBILE FINANCIAL FRAUD WITHIN KENYA. EPRA International Journal of Research and Development (IJRD), 8(1), pp.32-39.

[14] Xu, B., Wang, Y., Liao, X. and Wang, K., 2023. Efficient Fraud Detection using Deep Boosting Decision Trees. arXiv preprint arXiv:2302.05918.

[15] Fanai, H. and Abbasimehr, H., 2023. A novel combined approach based on deep Autoencoder and deep classifiers for credit card fraud detection. Expert Systems with Applications, p.119562.

[16] Xu, M., Fu, Y. and Tian, B., An ensemble fraud detection approach for online loans based on application usage patterns. Journal of Intelligent & Fuzzy Systems, (Preprint), pp.1-14.

[17] Alwadain, A., Ali, R.F. and Muneer, A., 2023. Estimating Financial Fraud through Transaction-Level Features and Machine Learning. Mathematics, 11(5), p.1184.

[18] Van Belle, R., Baesens, B. and De Weerdt, J., 2023. CATCHM: A novel network-based credit card fraud detection method using node representation learning. Decision Support Systems, 164, p.113866.

[19] Teng, H., Wang, C., Yang, Q., Chen, X. and Li, R., 2023. Leveraging Adversarial Augmentation on Imbalance Data for Online Trading Fraud Detection. IEEE Transactions on Computational Social Systems.

[20] Akinbowale, O.E., Mashigo, P. and Zerihun, M.F., 2023. The integration of forensic accounting and big data technology frameworks for internal fraud mitigation in the banking industry. Cogent Business & Management, 10(1), p.2163560.

[21] Zhou, H., Sun, G., Fu, S., Wang, L., Hu, J. and Gao, Y., 2021. Internet financial fraud detection based on a distributed big data approach with node2vec. IEEE Access, 9, pp.43378-43386.

[22] Varol, Altay, E.; Alatas, B. Bird swarm algorithms with chaotic mapping. Artif. Intell. Rev. 2020, 53, 1373–1414.

[23] Roy, S.; Chaudhuri, S.S. Cuckoo search algorithm using levy flight: A review. Int. J. Mod. Educ. Comput. Sci. 2013, 5, 10.