IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Revue d'Intelligence Artificielle
Vol. 34, No. 4, August, 2020, pp. 445-456
Journal homepage: http://iieta.org/journals/ria

Judgement Classification Using Hybrid ANN-Shuffled Frog Leaping Model on Cyber Crime Judgement Database

Radha Mothukuri[1], Bobba Basaveswararao[1], Suneetha Bulla[2*]

[1] Dept. of CSE, Acharya Nagarjuna University, Guntur 522510, India
[2] Dept. of CSE, Vigna's Foundation of Science, Technology and Research, Guntur, 522213, India

Corresponding Author Email: drsb_cse@vignan.ac.in

https://doi.org/10.18280/ria.340409

Received: 20 June 2020
Accepted: 5 August 2020

Keywords:
judgement case classification, shuffled frog leaping model, optimization

ABSTRACT

The world has taken dramatic transformation after advent of Information Technology, it is hard to find the people without cyber connected and every activity of us is guided and regulated by the connected networks. As the world is depending upon the information technology there is same extent of research is getting on cyber monitoring activities taking place around the world. Now, it is very vital to classify and prediction of cybercrimes on the connected era. The objective of the paper is to classify the cyber crime judgments precedents for providing knowledgeable and relevant information to the cyber crime legal stakeholders. The stakeholders extract information from the precedents is a crucial research problem because so much of judgments available in a digital form with remarkable evaluation of internet and bid data analytics. It is necessary to classify the precedents and to provide a bird- eye view of the relevant legal topics. In this study cybercrime related 2500 judgments are considered for evaluation of the Feed Forward Neural - Shuffled Frog Leaping (FNN-SFL) model. To achieve this objective a Feed Forward Neural based model with tuning of Term weights by adaption of a Bio Inspired tuning model Shuffled Frog Leaping model. The experiments are conducted and implemented the newly proposed FNN-SFL algorithm. The results and discussions are presented. The conclusions and future scope are presented at the end of the paper.

1. INTRODUCTION

Data Mining, just as information revelation, is the framework helped procedure of making by removal through and examining monstrous arrangements of information and afterward extricating the valuable data present in the information. It is additionally a strategy for finding astute, intriguing, and novel examples, just as enlightening, justifiable, and prescient models from enormous scale information [1]. In a basic manner, Data mining alludes to the way toward separating information that is important to the user. Due to the intense increment of advanced information the web, innovation allows the framework to play out the outline procedure to get to the abbreviated form of the computerized information [2]. Such innovation was executed in different fields to improve the advancement of work completed identified with that.

These days Legal Experts need the examination network to do some innovative creation to limit their work pressure and to accelerate the procedure. In this way the synopsis strategy was executed in the lawful field, to upgrade the judgment outline process. Indian Legal System follows the Statutes just as the Common Law.

Resolutions were the authoritative procedure or regulations issued by the Government, while Common Law was created by the judges through choices of courts and councils. In detail, a Common law is likewise called as 'Precedent', a standard of law which is built up by a court just because for a specific sort of case and after that it is alluded for basic leadership in comparable cases. Choices of the judges are the wellsprings of law. Right now, lawful experts were doing the complex administrative work of deciphering the lawful focuses and condensing the past judgment substance for their case contentions or to settle on the choice from them, such procedure needs precision and speed. Human-produced synopses need additional time and labour and are moderately costly. Creating the judgment outline is a repetitive undertaking too. In this way NLP based Summarization Techniques satisfy the requirements of the legitimate specialists in a basic and productive way. In this paper we have developed an effective FNN-SFL model to classify the documents based on their relevancy.

Artificial Neural Network in machine learning plays a vital role in classification and prediction. Feed forward Neural Network (FNN) is one of the sub sections of ANN where the flow of data will be in forward manner. There exists no backward propagation to tune the weights. In such cases the choice of weights goes random which may degrade the performance of FNN. In recent days evolutionary algorithms are used for solving these optimization problems. Shuffled Frog Leaping is one such optimization model works based on the memeplexes of frogs. In this paper the SFL algorithm is used for solving FNN regarding tuning of random weights.

Occasion based learning calculations are sluggish learning calculations [3], as they defer the enlistment or speculation process until characterization is performed. Sluggish learning calculations require less calculation time during the preparation stage than energetic learning calculations, (for example, choice trees, neural and Bayes nets) yet more calculation time during the arrangement procedure. One of the clearest occurrence-based learning calculations is the closest neighbour calculation. Zaki and Jr Meira [4] introduced an audit of occasion-based learning classifiers. Along these lines, right now, from a short portrayal of the nearest neighbour calculation, we will allude to some later works. K-Nearest Neighbour (KNN) depends on the rule that the occurrences inside a dataset will for the most part exist in nearness to different cases that have comparative properties [5]. In the event that the occurrences are labelled with an arrangement mark, at that point the estimation of the name of an unclassified occasion can be dictated by watching the class of its closest neighbours. The KNN finds the k closest examples to the question case and decides its class by recognizing the absolute most continuous class label.

A study of weighting plans is given by Wettschereck et al. [6]. The intensity of KNN has been exhibited in various genuine spaces, however there are a few hesitations about the handiness of KNN, for example, I) they have huge stockpiling prerequisites, ii) they are delicate to the decision of the likeness work that is utilized to think about cases, iii) they do not have a principled method to pick k, aside from through cross-approval or comparable, computationally-costly procedure [7]. The decision of k influences the exhibition of the KNN calculation. Consider the accompanying reasons why a K-Nearest Neighbour classifier may mistakenly arrange an inquiry occasion:

Wettschereck et al. [6] explored the conduct of the KNN within the sight of loud occurrences. The trials indicated that the presentation of KNN was not delicate to the specific decision of k when k was huge. They found that for little estimations of k, the KNN calculation was stronger than the single closest neighbour calculation (1NN) for most of enormous datasets tried. Be that as it may, the exhibition of the KNN was mediocre compared to that accomplished by the 1NN on little datasets.

Okamoto and Yugami [8] spoke to the normal grouping exactness of k-NN as an element of area qualities including the quantity of preparing occasions, the quantity of significant and unessential properties, the likelihood of each characteristic, the commotion rate for each kind of clamor, and k. They likewise investigated the conduct ramifications of the examinations by exhibiting the impacts of area qualities on the normal precision of k-NN and on the ideal estimation of k for fake areas.

An opportunity to group the question occasion is firmly identified with the quantity of put away occurrences and the quantity of highlights that are utilized to portray each example. Along these lines, so as to decrease the quantity of put away occurrences, case sifting calculations have been proposed by Kubat and Cooperson [9]. Brighton and Mellish [10] found that their ICF calculation and RT3 calculation [11] accomplished the most noteworthy level of example set decrease just as the maintenance of characterization precision: they are near accomplishing unintrusive stockpiling decrease. How much these calculations perform is very great: a normal of 80% of cases is evacuated and grouping precision doesn't drop essentially. One other decision in planning preparing set decrease calculation is to change the occasions utilizing another portrayal, for example, models [12] revealed that the solidness of closest neighbour classifiers recognizes them from choice trees and a few sorts of neural systems. A learning technique is named "temperamental" if little changes in the preparation test set split can bring about huge changes in the subsequent classifier. As we have just referenced, the significant weakness of occasion-based classifiers is their enormous computational time for arrangement. A key issue in numerous applications is to figure out which of the accessible info highlights ought to be utilized in displaying by means of highlight determination [13], on the grounds that it could improve the arrangement precision and scale down the necessary characterization time. Moreover, picking an increasingly reasonable separation metric for the particular dataset can improve the exactness of case-based classifiers.

The novel contribution of paper is to classify the cyber crime judgments precedents for providing expert and pertinent information to the cyber crime legal stakeholders. The stakeholders extract information from the precedents is a crucial research problem because so much of judgments available in a digital form with remarkable evaluation of internet and bid data analytics. It is necessary to classify the precedents and to provide a bird- eye view of the relevant legal topics. In this study cybercrime related 2500 judgments are considered for evaluation of the Feed Forward Neural - Shuffled Frog Leaping (FNN-SFL) model. To achieve this objective a Feed Forward Neural based model with tuning of Term weights by adaption of a Bio Inspired tuning model Shuffled Frog Leaping model.

2. PROBLEM DEFINITION

2.1 Classification of judgement database

The summarization of documents in a criminal case is the base for judgement classification. As a base model, the public prosecutors are the people who are responsible for examining the prosecution against the offender. Few are of with less impact and the judgement in the court will be summarised in the judgement case files. This will be done after the procedures of policemen being carried out. The policemen will be collecting the evidences of the crime scenes and the summary will be submitted in front of the judges. These are all the criterion that are carried out in hard criminal cases [14].

In terms of cyber crime cases the police will be submitting the proof's in the form of softcopy and they are not meant to be summarised in full fledged manner in the case files. In such cases the classification model is of high with difficulty to categorise into proforma.

In our model we used ANN model for classification purpose and the classification model will be taken care with the other procedures such as classification and weight tuning. In weight tuning we have incorporated Shuffled Frog Leaping algorithm for examining purposes.

2.2 Working principle multi-layer perceptron in FNN

An example module of Multilayer perceptron in Feed Forward Neural Network is given in Figure 1. It consists of 1 input layer which can take n number of input features and one hidden layer where the number of neurons can be higher or lessen than the input neurons and one other output layer where it has to be at least one.

The MLP output can be calculated as follows: Initially the weighted sum of inputs is computed using Eq. (1)

$$s_j = \sum_{i=1}^{n} \left(W_{i,j} \times X_i\right) - \theta_j \tag{1}$$

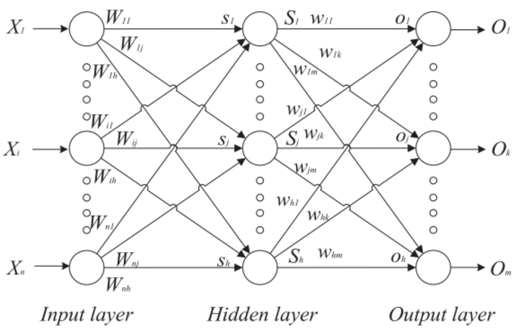

Figure 1. MLP

Such that the j ranges from 1 to n, and n represents the number of input neurons in the model. S refers the out of input neuron with the computation of it along with the weights and it will be the input for the hidden layer. The output of the hidden layer will be 3 computed as follows

$$S_j = \frac{1}{(1 + \exp(-s_j))} \quad (2)$$

where, j ranges from 1 to h such that h in the number of hidden neurons in the hidden layer.

Then the final output can be computed as

$$o_k = \sum_{j=1}^{h} (W_{j,k} \times S_j) - \theta_k \quad (3)$$

where, k ranges from 1 to m.

$$O_k = \frac{1}{(1 + \exp(-o_k))} \quad (4)$$

where, $W_{j,k}$ represents the weight of each edge from hidden layer j to the output layer node k, θ_k is the bias value of output layer node k.

In MLP the weights and bias values have a great impact on outputs. When the weights and bias values are tuned to obtain the ideal output values then the classification of test datasets will be ideal towards the predicted output. Training of MLP includes obtaining optimal values for each value of weight and bias.

2.3 Tuning of weight parameters

The weights and the bias values are the most prominent features of FNN which possess the theme to produce better classification model of the given problem. In our cases the problem will be described as a set of features for every judgement case file. Each file will be represented as a set of features and each feature in the judgement case will act as an input to FNN. The weights in FNN tunes itself to an optimal value using the mathematical equations so that the output classification to be crisp and effective. In this paper we proposed Shuffled Frog Leaping algorithm for tuning the weights of FNN.

3. SFL ON TUNING WEIGHTS OF MLP

Basically, there are three different approaches are followed to train an MLP using heuristic algorithm.

(1): In the first approach, the heuristic algorithms are employed to obtain an optimal combination of weights and bias values which reduces the total minimal error rate.

(2): In the second approach, the MLP architecture will be designed using a heuristic algorithm w.r.t. the problem domain.

(3): The third approach is to employ the heuristic algorithm to fine tune the parameters of gradient based algorithms which further carry over the process of MLP.

In our method we follow the first method namely Vector based approach. An example of solution representation is given in Figure 2

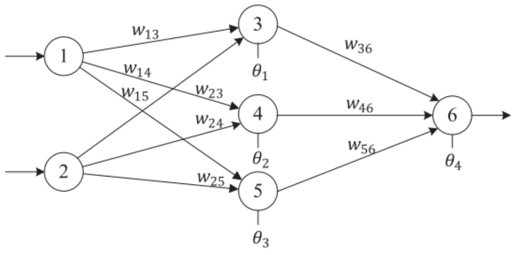

Figure 2. Structure of MLP (2-3-1)

2-3-1 indicates that the MLP in Figure 2 consists of two input nodes, three hidden nodes and one output node. The weights in this which are in need to be optimized includes $\{w_{1,3}, w_{1,4}, w_{1,5}, w_{2,3}, w_{2,4}, w_{2,5}, w_{3,6}, w_{4,6}, w_{5,6}\}$.

SHUFFLED FROG LEAPING ALGORITHM for Tuning the Weights in FNN

Variables Used

n← # number of weights
G←Maximum No. of Generations
$PopSize$←Total number of frogs
$f()$← objective function
p← Total number of memeplexes
q← number of frogs in each memeplex
$Pop_{Popsize \times n}$←Possible solutions / Search space
$Frog_{n \times d}$←Single solution in the possible solutions / Search space
D_{max}←Max allowed position drift by a individual
$Frog_v$←Virtual individual
Ub←upper bound (1)
Lb←lower bound (0)

ALGORITHM:

Step 1: Initialize$t \leftarrow 1, p, q, PopSize, G$
Step 2: Generate random values for frogs
For each $i = 1{:}PopSize$ do
$Frog_i \leftarrow ub + (ub - lb) * rand()$ *// rand is used to generate the random numbers between 0 to 1*
End
Step 3: Calculate the fitness using FNN
For each $i = 1{:}PopSize$ do
$Fit\_Frog_i \leftarrow f(Frog_i)$ *// f() is the objective function where the individuals are evaluated based on the fitness value. this fitness value will be used for ranking the solutions*
End
Step 4: Sort the frogs in ascending order based on the objective function *//Rank the frogs based on the*

fitness value. Sor the individuals in ascending order based on the fitness array.
Step 5: Partitioning the frogs into memeplexes
For each $i = 1{:}\,PopSize$ do
$memeplex_q \leftarrow imod(p)$ // *q ranges from 1 to q which indicates the total number of memeplexes.*
End
Step 6: for each memeplex Repeat through Step 10 **Until** $t \leq G$,
else go to Step 10
Step 7: iteration starts for each memeplex
Step 7.1: Re-frame the labels of each frogs with the memeplex number and the index of the frog in each memeplex.
Step 7.2: Compute the probability for each frog in the memeplex
$P(Frog_j) = 2(q+1-j)/q(q+1)$
Step 7.3: Generate a sub-memeplex with q frogs based on the random probability.
Step 7.3.1: Find the worst and the best frogs $Frog_B$ and $Frog_w$ based on the fitness value in each sub memeplex.
Step 7.3.2: Alter the position of the worst frog using the Equation
$T_j = rand() \times (Frog_B - Frog_w)$
Step 7.3.3: Add the new position into the existing frog position.
$Frog_{w_{new}} = Frog_w + T_j$
Step 7.3.4: Find the fitness of the frog.
$Fit\_Frog_{w_{new}} \leftarrow f(Frog_{w_{new}})$
Step 7.3.5: If $(Fit\_Frog_{w_{new}} < Fit\_Frog_w)$ // *for min problem*
$Frog_w = Frog_{w_{new}}$
Step 7.3.6: Else
Step 7.3.7: Alter the position of the worst frog using the Equation
$T_j = rand() \times (Frog_G - Frog_w)$
Step 7.3.8: Add the new position into the existing frog position.
$Frog_{w_{new}} = Frog_w + T_j$
Step 7.3.9: Find the fitness of the frog using FNN
$Fit\_Frog_{w_{new}} \leftarrow f(Frog_{w_{new}})$
Step 7.3.10: If $(Fit\_Frog_{w_{new}} < Fit\_Frog_w)$ // *for min problem*
$Frog_w = Frog_{w_{new}}$
Step 7.3.11: Else
Step 7.3.12: $Frog_w \leftarrow ub + (ub - lb) * rand()$
Step 8: Shuffle all the memeplexes // *Shuffle the groups once the number of sub iterations completed*
Step 9: $t = t + 1$
Step 10: Find the final best solution stored in $Frog_b$

OUTPUT: $Frog_B$

4. EXPERIMENTAL ANALYSIS

This section includes the subsections such as Experimental setup for classification, performance metrics to prove the significance of the proposed approach, the analysis of the results.

4.1 Experimental setup

The proposed model is developed in MATLAB 2015a version in the system with specifications of Processor as Core i7 3rd Generation with a clock speed of 3.2 GHz, 4GB DDR4 RAM and 1TB HDD. For examining the significance of the proposed model has been done over 2500 documents from Supreme Court of India. For classification purpose we have taken 759 of child pornography related documents and 574 grooming related case documents from 1333 documents (Table 1).

Table 1. Parameter settings

Parameters	Range
Initial Population	100
Maximum Iterations	1000
P_a	0.25
Runs	20
D	Based on FNN model

4.2 Performance metrics

Based on the confusion matrix (Table 2) the performance of the proposed model and the existing model are evaluated.

Table 2. Confusion matrix and notations

		True Condition	
	Total Population	Condition Positive (P)	Condition Negative (N)
Predicted Condition	Prediction Condition Positive	True Positive (TP)	False Positive (FP)
	Prediction Condition Negative	False Negative (FN)	True Negative (TN)

The following performance metrics are used to evaluate the proposed schema in terms of accuracy and false prediction rate. The performance metrics are defined as follows:

4.2.1 Accuracy

Accuracy is defined as the ration between the sum of true positive as well as true negative and the total number of sample models including conditional positive and negative.

$$Accuracy = \frac{TP + TN}{P + N} \tag{5}$$

4.2.2 Precision

Precision is defined as the ratio between the true positive values and the sum of true positive and false positive values.

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

4.2.3 Sensitivity (Recall)

Sensitivity is defined as the ratio between the true positive values and the conditional positive values. It can be represented as

$$Sensitivity = \frac{TP}{P} \tag{7}$$

4.2.4 Specificity

Specificity is defined as the ratio between the total number of true negative values to the conditional negative values. It can be represented as

$$Specificity = \frac{TN}{N} \quad (8)$$

4.2.5 Miss rate

Miss rate is defined as the ratio between the total number of false negative values to the conditional positive values. It can be represented as

$$MissRate = \frac{FN}{P} \quad (9)$$

4.3 Result analysis

4.3.1 Result Analysis w.r.t. Accuracy

In Table 3 the comparison model of FNN-SFL with existing methods w.r.t. Accuracy on classification of Child Pornography Documents has been given. This table shows the significance of FNN-SFL on Accuracy over other existing models.

According to the Figure 3 respect to accuracy on classification FNN-SFL outperforms the existing models along with the feature selection model Eff. The significance in terms of percentage against KNN with 5.71%, against Decision Tree with 1.25%, against Gaussian Naïve Bayes with 2.5%, against Logistic Regression with 3.21%, against Random Forest with 1.62% and against SVM with 4.57%. With KS Model FNN-SFL outperforms existing models such as KNN with 3.19%, against Decision Tree with 0.49%, against Gaussian Naïve Bayes with 5.05%, against Logistic Regression with 3.22%, against Random Forest with 2.74% and against SVM with 7.01%. With Correlation Model FNN-SFL outperforms existing models such as KNN with 3.60%, against Decision Tree with 1.75%, against Gaussian Naïve Bayes with 0.36%, against Logistic Regression with 1.42%, against Random Forest with 4.66% and against SVM with 6.84%.

With BSFL, FNN-SFL outperforms existing models such as KNN with 9.43%, against Decision Tree with 5.51%, against Gaussian Naïve Bayes with 7.09%, against Logistic Regression with 4.34%, against Random Forest with 1.87% and against SVM with 3.63%.

In Table 4 the comparison model of FNN-SFL with existing methods w.r.t. Accuracy on classification of Grooming Documents has been given. This table shows the significance of FNN-SFL on Accuracy over other existing models.

In Detail with based on the Figure 4 respect to accuracy on classification FNN-SFL outperforms the existing models along with the feature selection model Eff. The significance in terms of percentage against KNN with 4.89%, against Decision Tree with 0.24%, against Gaussian Naïve Bayes with 3.5%, against Logistic Regression with 5.42%, against Random Forest with 3.69%, against FNN with 0.73% and against SVM with 0.04%. With KS Model FNN-SFL outperforms existing models such as KNN with 0.19%, against Decision Tree with 4.97%, against Gaussian Naïve Bayes with 2.52%, against Logistic Regression with 9.09%, against Random Forest with 3.94%, against FNN with 0.93% and against SVM with 4.45%. With Correlation Model FNN-SFL outperforms existing models such as KNN with 2.19%, against Decision Tree with 3.74%, against Gaussian Naïve Bayes with 6.36%, against Logistic Regression with 9.48%, against Random Forest with 5.12%, against FNN with 1.96% and against SVM with 7.49%.

With BSFL, FNN-SFL outperforms existing models such as KNN with 5.26%, against Decision Tree with 4.19%, against Gaussian Naïve Bayes with 6.13%, against Logistic Regression with 4.69%, against Random Forest with 5.36% and against SVM with 9.14%.

Table 3. Comparison of FNN-SFL with existing methods w.r.t. accuracy on classification of child pornography documents

Accuracy	KNN [13]	Decision Tree [15]	Gaussian Naïve Bayes [16]	Logistic Regression [17]	Random Forest [18]	SVM [19]	FNN [14]	FNN-SFL
Eff	86.57	90.67	89.52	88.87	90.33	87.62	84.5	91.82
KS	82.06	84.35	80.49	82.04	82.44	78.83	71.68	84.77
Correlation	87.86	89.54	90.81	89.84	86.89	84.91	82.41	91.14
BSFL	84.79	88.46	86.98	89.56	91.87	90.22	85.43	93.62

Table 4. Comparison of FNN-SFL with existing methods w.r.t. accuracy on classification of grooming documents

Accuracy	KNN	Decision Tree	Gaussian Naïve Bayes	Logistic Regression	Random Forest	SVM	FNN	FNN-SFL
Eff	86.80	91.04	88.07	86.32	87.89	91.23	90.43	91.27
KS	86.02	81.90	84.01	78.34	82.78	82.35	79.31	86.18
Correlation	88.26	86.85	84.49	81.67	85.61	83.47	82.05	90.23
BSFL	88.70	89.70	87.88	89.23	88.60	85.06	88.22	93.62

Table 5. Comparison of FNN-SFL with existing methods w.r.t. precision on classification of child pornography documents

Precision	KNN	Decision Tree	Gaussian Naïve Bayes	Logistic Regression	Random Forest	SVM	FNN	FNN-SFL
Eff	83.79	85.36	88.43	88.16	85.29	87.06	85.35	88.84
KS	83.57	85.77	84.22	88.37	89.27	87.76	87.77	89.65
Correlation	85.72	88.94	88.70	90.33	91.23	84.76	82.36	91.99
BSFL	87.98	88.63	91.08	88.56	89.45	90.83	90.01	92.65

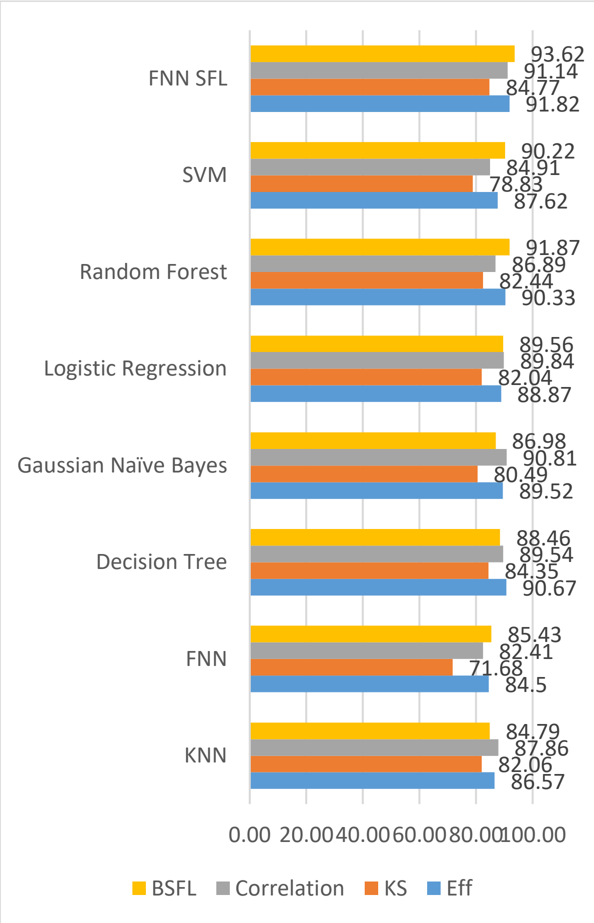

Figure 3. Comparison graph of FNN-SFL with existing methods w.r.t. accuracy on classification of child pornography documents

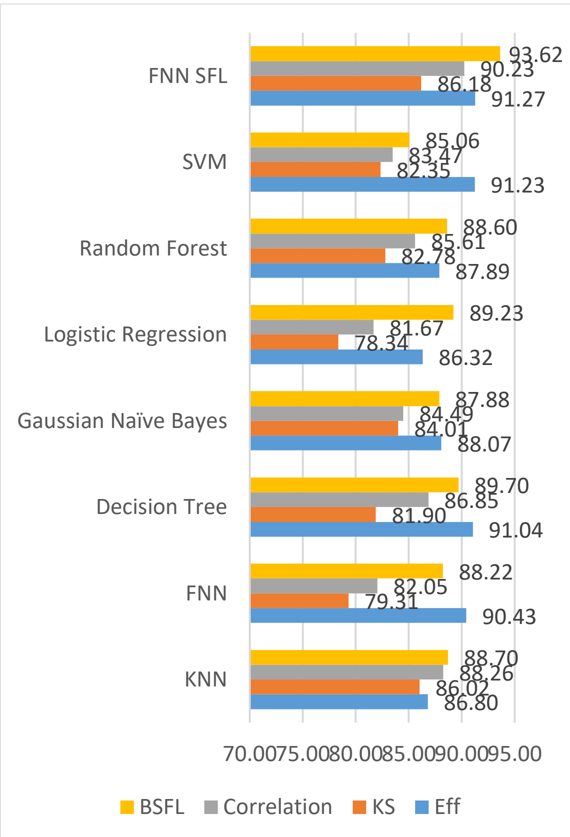

Figure 4. Comparison graph of FNN-SFL with existing methods w.r.t. accuracy on classification of grooming documents

4.3.2 Result analysis w.r.t. precision

In Table 5 the comparison model of FNN-SFL with existing methods w.r.t. Precision on classification of Child Pornography Documents has been given. This table shows the significance of FNN-SFL on precision over other existing models.

Based on the Figure 5 in Detail with respect to precision on classification FNN-SFL outperforms the existing models along with the feature selection model Eff. The significance in terms of percentage against KNN with 5.68%, against Decision Tree with 3.91%, against Gaussian Naïve Bayes with 0.45%, against Logistic Regression with 0.76%, against Random Forest with 3.99%, against FNN with 2.85% and against SVM with 2.01%. With KS Model FNN-SFL outperforms existing models such as KNN with 6.78%, against Decision Tree with 4.32%, against Gaussian Naïve Bayes with 6.06%, against Logistic Regression with 1.42%, against Random Forest with 0.42%, against FNN with 6.24%and against SVM with 2.11%. With Correlation Model FNN-SFL outperforms existing models such as KNN with 6.81%, against Decision Tree with 3.30%, against Gaussian Naïve Bayes with 3.57%, against Logistic Regression with 1.79%, against Random Forest with 0.82%, against FNN with 6.41%and against SVM with 7.85%.

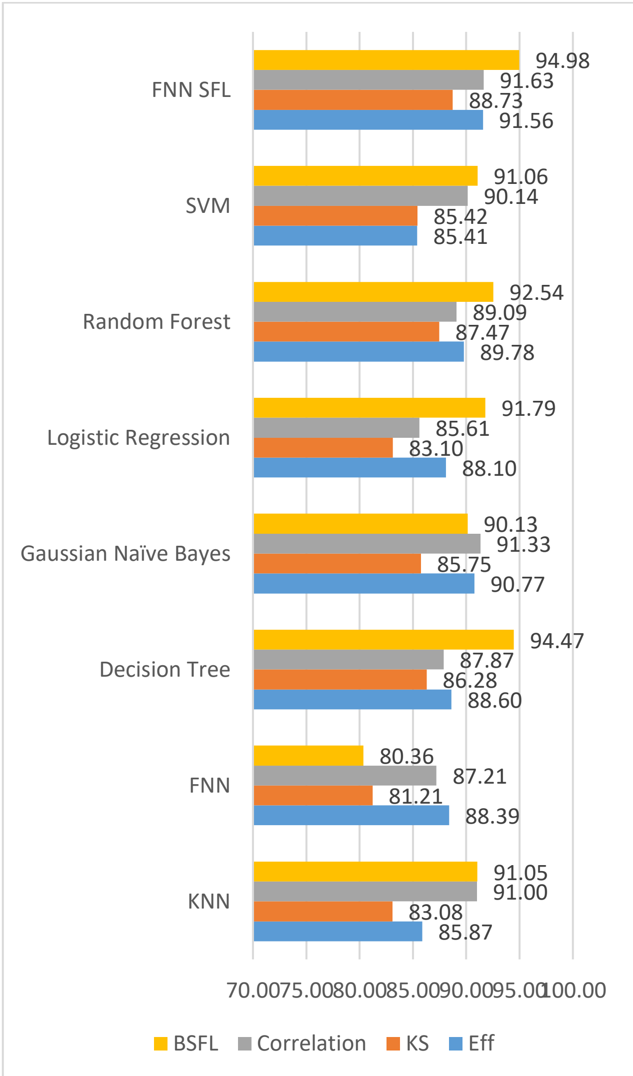

Figure 5. Comparison graph of FNN-SFL with existing methods w.r.t. precision on classification of child pornography documents

Table 6. Comparison of FNN-SFL with existing methods w.r.t. precision on classification of grooming documents

Precision	KNN	Decision Tree	Gaussian Naïve Bayes	Logistic Regression	Random Forest	SVM	FNN	FNN-SFL
Eff	85.87	88.60	90.77	88.10	89.78	85.41	88.39	91.56
KS	83.08	86.28	85.75	83.10	87.47	85.42	81.21	88.73
Correlation	91.00	87.87	91.33	85.61	89.09	90.14	87.21	91.63
BSFL	91.05	94.47	90.13	91.79	92.54	91.06	80.36	94.98

Table 7. Comparison of FNN - SFL with existing methods w.r.t. sensitivity on classification of child pornography documents

Sensitivity	KNN	Decision Tree	Gaussian Naïve Bayes	Logistic Regression	Random Forest	SVM	FNN	FNN-SFL
Eff	92.43	91.93	89.72	90.79	92.52	90.08	88.95	94.19
KS	90.02	93.99	90.75	86.21	91.32	93.01	90.65	94.42
Correlation	90.95	92.62	90.58	85.78	91.42	89.29	90.58	94.37
BSFL	91.93	93.66	94.02	91.17	89.09	92.82	91.54	94.60

Table 8. Comparison of FNN-SFL with existing methods w.r.t. sensitivity on classification of grooming documents

Sensitivity	KNN	Decision Tree	Gaussian Naïve Bayes	Logistic Regression	Random Forest	SVM	FNN	FNN-SFL
Eff	84.51	85.36	82.01	87.06	87.70	88.01	80.64	89.59
KS	81.92	85.26	86.06	85.89	84.58	85.23	80.59	88.48
Correlation	86.51	81.46	80.23	85.15	83.50	84.29	81.45	88.33
BSFL	86.19	90.53	87.60	84.73	88.75	86.17	81.76	92.33

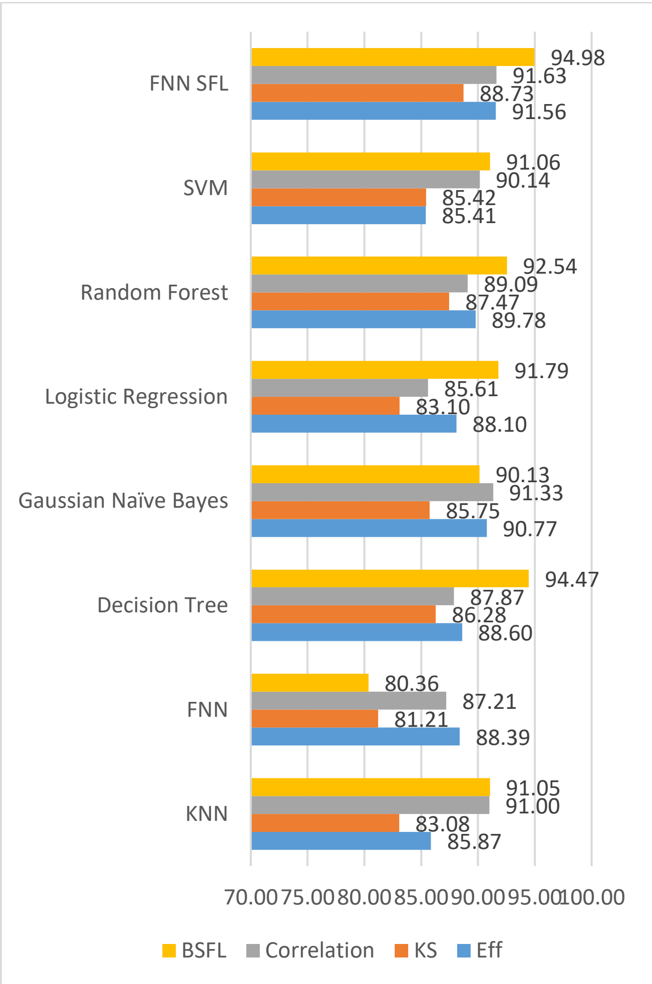

Figure 6. Comparison graph of FNN-SFL with existing methods w.r.t. precision on classification of grooming documents

With BSFL, FNN-SFL outperforms existing models such as KNN with 5.03%, against Decision Tree with 4.32%, against Gaussian Naïve Bayes with 1.69%, against Logistic Regression with 4.41%, against Random Forest with 3.45% and against SVM with 1.96%.

In Table 6 the comparison model of FNN-SFL with existing methods w.r.t. Precision on classification of Grooming Documents has been given. This table shows the significance of FNN-SFL on precision over other existing models.

In Detail with based on the Figure 6 respect to precision on classification FNN-SFL outperforms the existing models along with the feature selection model Eff. The significance in terms of percentage against KNN with 6.22%, against Decision Tree with 3.23%, against Gaussian Naïve Bayes with 0.86%, against Logistic Regression with 3.78%, against Random Forest with 1.94%, against FNN with 4.57% and against SVM with 6.72%. With KS Model FNN-SFL outperforms existing models such as KNN with 6.37%, against Decision Tree with 2.76%, against Gaussian Naïve Bayes with 3.36%, against Logistic Regression with 6.34%, against Random Forest with 1.42%, against FNN with 4.62% and against SVM with 3.73%. With Correlation Model FNN-SFL outperforms existing models such as KNN with 0.68%, against Decision Tree with 4.1%, against Gaussian Naïve Bayes with 0.32%, against Logistic Regression with 6.57%, against Random Forest with 2.77%, against FNN with 7.62% and against SVM with 1.62%.

With BSFL, FNN-SFL outperforms existing models such as KNN with 4.14%, against Decision Tree with 0.54%, against Gaussian Naïve Bayes with 5.11%, against Logistic Regression with 3.35%, against Random Forest with 2.57% and against SVM with 4.13%.

4.3.3 Result analysis w.r.t. sensitivity

In Table 7 the comparison model of FNN-SFL with existing methods w.r.t. sensitivity on classification of Child Pornography Documents has been given. This table shows the significance of FNN-SFL on sensitivity over other existing models.

As per the Figure 7 with BSFL, FNN-SFL outperforms existing models such as KNN with 2.81%, against Decision Tree with 0.99%, against Gaussian Naïve Bayes with 0.61%, against Logistic Regression with 3.62%, against Random Forest with 5.82% and against SVM with 1.88%.

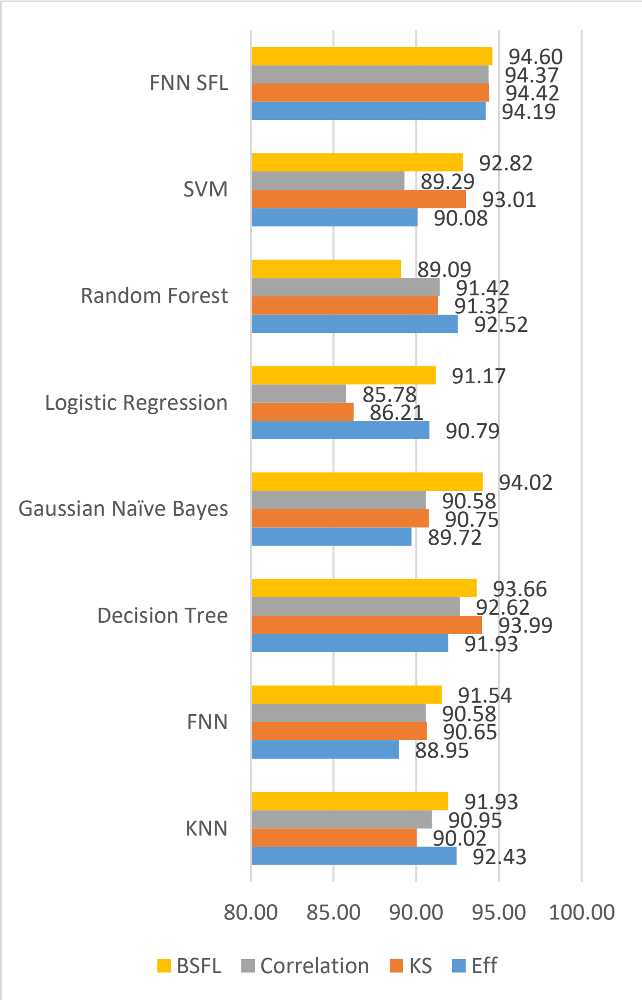

Figure 7. Comparison graph of FNN-SFL with existing methods w.r.t. sensitivity on classification of child pornography documents

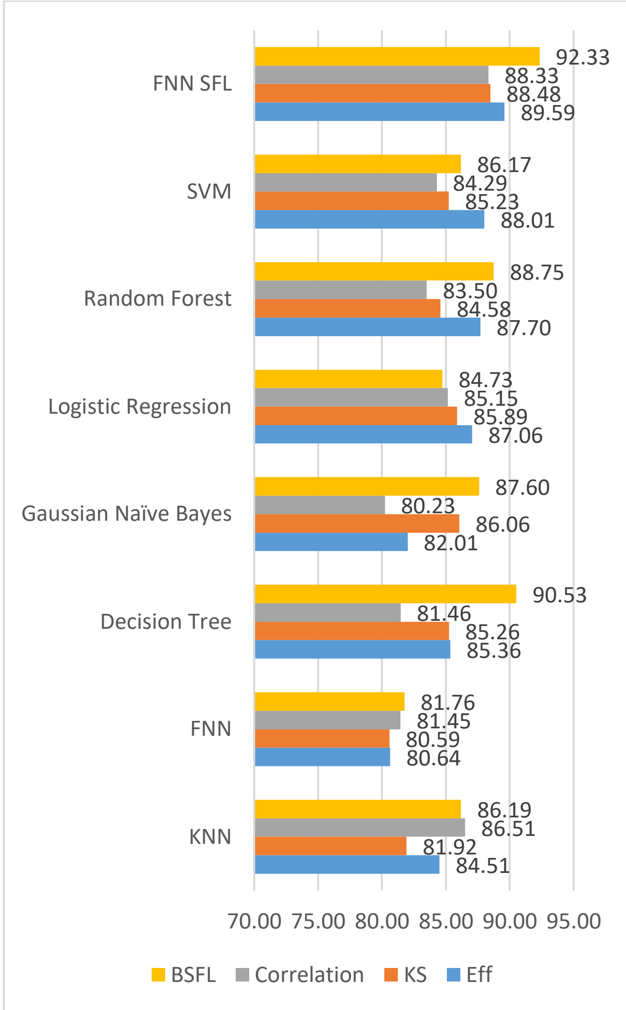

Figure 8. Comparison graph of FNN-SFL with existing methods w.r.t. sensitivity on classification of grooming documents

In Table 8 the comparison model of FNN-SFL with existing methods w.r.t. sensitivity on classification of grooming Documents has been given. This table shows the significance of FNN-SFL on sensitivity over other existing models

The Figure 8 depicts to Sensitivity on classification FNN-SFL outperforms the existing models along with the feature selection model Eff. The significance in terms of percentage against KNN with 5.67%, against Decision Tree with 4.72%, against Gaussian Naïve Bayes with 8.46%, against Logistic Regression with 2.82%, against Random Forest with 2.11%, against FNN with 4.93%and against SVM with 1.76%. With KS Model FNN-SFL outperforms existing models such as KNN with 7.40%, against Decision Tree with 3.63%, against Gaussian Naïve Bayes with 2.72%, against Logistic Regression with 2.92%, against Random Forest with 4.40%, against FNN with 9.62% and against SVM with 3.66%. With Correlation Model FNN-SFL outperforms existing models such as KNN with 2.06%, against Decision Tree with 7.77%, against Gaussian Naïve Bayes with 9.17%, against Logistic Regression with 3.60%, against Random Forest with 5.46%, against FNN with 4.12% and against SVM with 4.58%.

With BSFL, FNN-SFL outperforms existing models such as KNN with 6.65%, against Decision Tree with 1.95%, against Gaussian Naïve Bayes with 5.13%, against Logistic Regression with 8.23%, against Random Forest with 3.88% and against SVM with 6.67%.

4.3.4 Result analysis w.r.t. specificity

Table 9 shows the comparison model of SFL-ANN with existing methods w.r.t. specificity on classification of child pornography Documents has been given. This table shows the significance of SFL-ANN on specificity over other existing models.

The Figure 9 shows specificity on classification FNN-SFL outperforms the existing models along with the feature selection model Eff. The significance in terms of percentage against KNN with 0.96%, against Decision Tree with 3.19%, against Gaussian Naïve Bayes with 0.83%, against Logistic Regression with 2.78%, against Random Forest with 5.38%, against FNN with 5.43% and against SVM with 3.48%. With KS Model FNN-SFL outperforms existing models such as KNN with 5.87%, against Decision Tree with 3.64%, against Gaussian Naïve Bayes with 6.77%, against Logistic Regression with 1.27%, against Random Forest with 2.35%, against FNN with 9.24% and against SVM with 1.16%. With Correlation Model FNN-SFL outperforms existing models such as KNN with 0.84%, against Decision Tree with 3.25%, against Gaussian Naïve Bayes with 6.67%, against Logistic Regression with 0.56%, against Random Forest with 4.15%, against FNN with 7.35% and against SVM with 0.77%.

Table 10 shows the comparison of FNN-SFL with existing methods w.r.t. specificity on classification of grooming documents.

Table 9. Comparison of SFL-ANN with existing methods w.r.t. specificity on classification of child pornography documents

Specificity	KNN	Decision Tree	Gaussian Naïve Bayes	Logistic Regression	Random Forest	SVM	FNN	FNN-SFL
Eff	87.20	84.61	87.32	85.61	83.31	84.98	80.66	88.06
KS	86.24	88.29	85.43	90.46	89.47	90.57	81.52	91.63
Correlation	85.67	83.59	80.64	85.92	82.81	85.74	75.84	86.41
BSFL	86.89	87.62	90.97	86.21	90.71	85.09	79.65	92.33

Table 10. Comparison of FNN-SFL with existing methods w.r.t. specificity on classification of grooming documents

Specificity	KNN	Decision Tree	Gaussian Naïve Bayes	Logistic Regression	Random Forest	SVM	FNN	FNN-SFL
Eff	88.25	84.80	85.34	87.57	83.88	83.74	84.74	90.10
KS	85.69	89.13	87.22	82.36	86.27	87.17	80.69	89.54
Correlation	82.14	88.05	85.21	88.73	88.69	82.60	75.22	89.36
BSFL	89.03	93.64	93.33	86.42	92.41	92.33	82.54	94.60

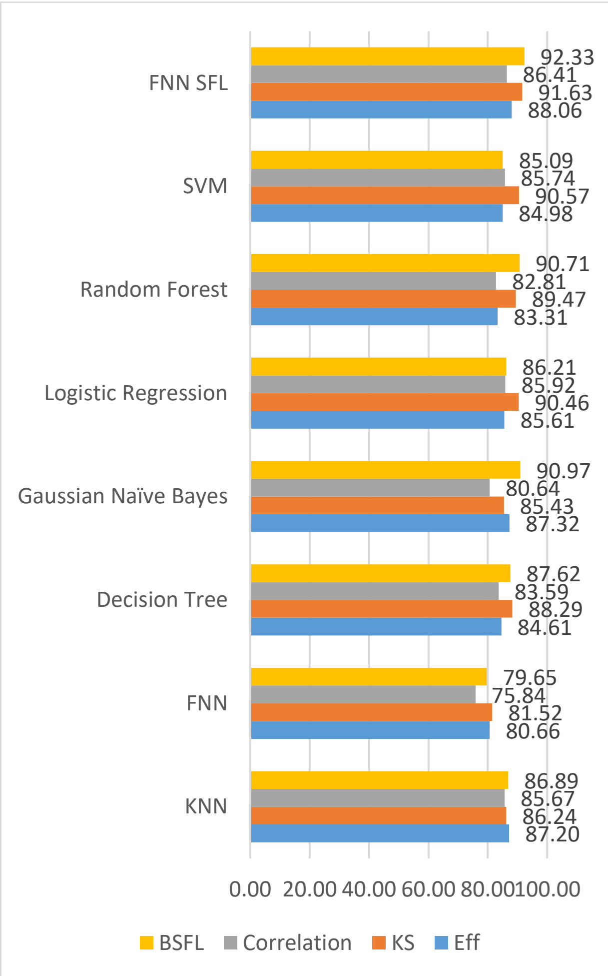

Figure 9. Comparison graph of FNN-SFL with existing methods w.r.t. specificity on classification of child pornography documents

According to Figure 10 respect to Specificity on classification FNN-SFL outperforms the existing models along with the feature selection model Eff. The significance in terms of percentage against KNN with 2.05%, against Decision Tree with 5.89%, against Gaussian Naïve Bayes with 5.28%, against Logistic Regression with 2.80%, against Random Forest with 6.90%, against FNN with 7.62% and against SVM with 7.06%. With KS Model FNN-SFL outperforms existing models such as KNN with 4.29%, against Decision Tree with 0.45%, against Gaussian Naïve Bayes with 2.59%, against Logistic Regression with 8.01%, against Random Forest with 3.65%, against FNN with 9.52% and against SVM with 2.64%. With Correlation Model FNN-SFL outperforms existing models such as KNN with 8.08%, against Decision Tree with 1.47%, against Gaussian Naïve Bayes with 4.64%, against Logistic Regression with 0.71%, against Random Forest with 0.74%, against FNN with 9.54% and against SVM with 7.56%.

With BSFL, FNN-SFL outperforms existing models such as KNN with 5.88%, against Decision Tree with 1.01%, against Gaussian Naïve Bayes with 1.34%, against Logistic Regression with 8.64%, against Random Forest with 2.31% and against SVM with 2.40%.

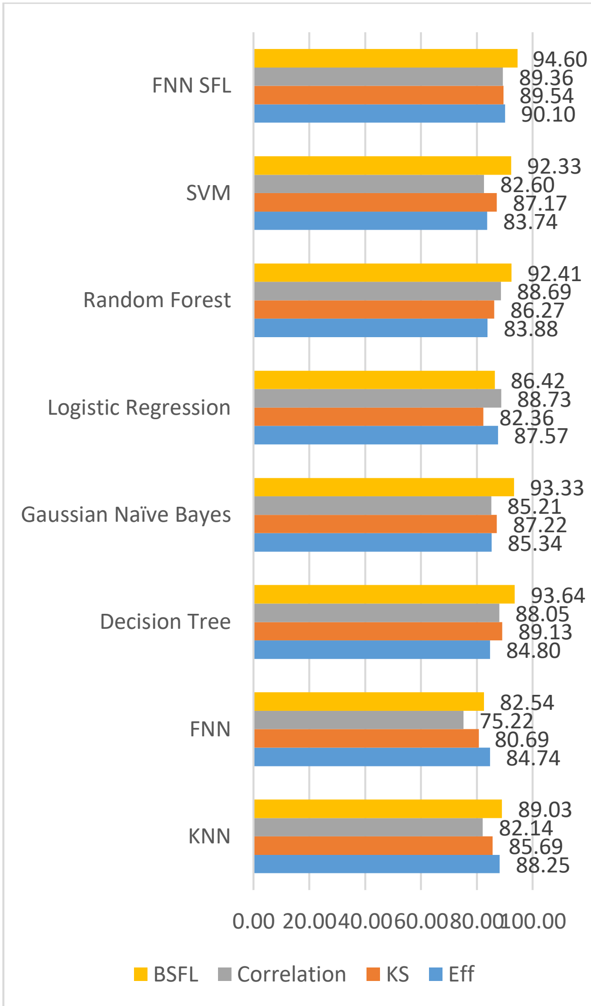

Figure 10. Comparison graph of FNN-SFL with existing methods w.r.t. specificity on classification of grooming documents

Table 11. Comparison of FNN-SFL with existing methods w.r.t. Miss ratio on classification of child pornography documents

Miss Ratio	KNN	Decision Tree	Gaussian Naïve Bayes	Logistic Regression	Random Forest	SVM	FNN	FNN-SFL
Eff	4.19	4.11	4.15	4.00	4.23	4.05	4.01	3.98
KS	3.78	3.85	3.79	3.86	3.74	3.84	4.87	3.72
Correlation	3.93	3.94	4.03	4.11	4.04	4.04	3.99	3.91
BSFL	3.76	3.70	3.82	3.69	3.88	3.80	3.78	3.69

Table 12. Comparison of FNN-SFL with existing methods w.r.t. Miss ratio on classification of grooming documents

Miss Ratio	KNN	Decision Tree	Gaussian Naïve Bayes	Logistic Regression	Random Forest	SVM	FNN	FNN-SFL
Eff	10.56	10.68	10.69	10.51	10.54	10.59	10.66	10.17
KS	10.58	10.66	10.38	10.82	11.04	10.74	10.52	10.20
Correlation	10.45	10.33	10.18	10.67	10.50	10.32	10.21	10.14
BSFL	9.94	10.20	10.86	9.98	10.61	10.36	9.99	9.93

4.3.5 Result analysis w.r.t. Miss ratio

In Table 11 the comparison model of FNN-SFL with existing methods w.r.t. Miss ratio on classification of child pornography Documents has been given. This table shows the significance of FNN-SFL on miss ratio over other existing models.

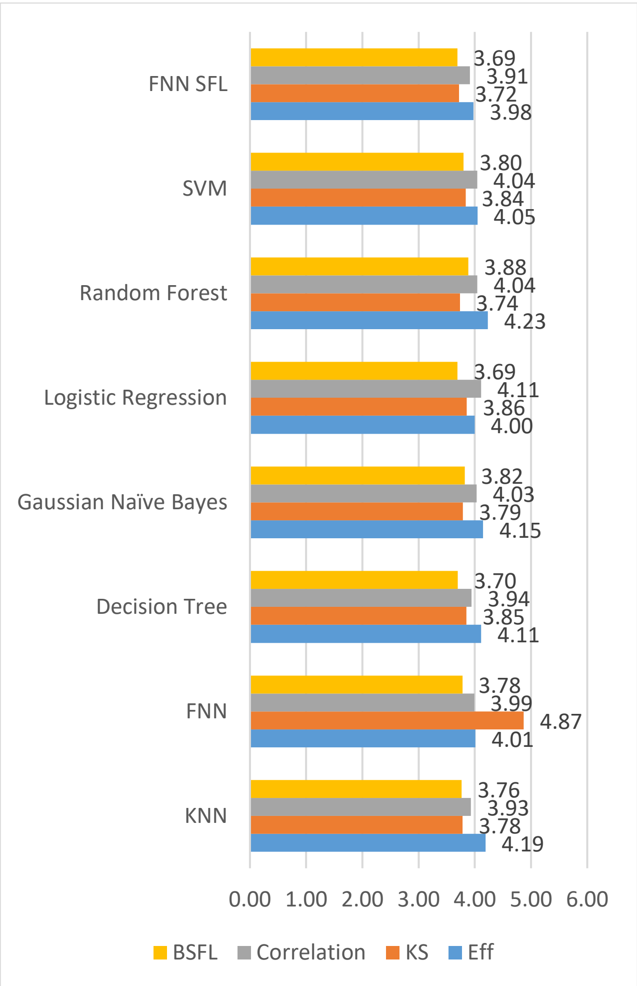

Figure 11. Comparison graph of FNN-SFL with existing methods w.r.t. Miss ratio on classification of child pornography documents

Figure 11 shows the specificity on classification FNN-SFL outperforms the existing models along with the feature selection model Eff. The significance in terms of percentage against KNN with 5.07%, against Decision Tree with 3.31%, against Gaussian Naïve Bayes with 4.12%, against Logistic Regression with 0.48%, against Random Forest with 6.02%, against FNN with 2.43% and against SVM with 1.74%. With KS Model FNN-SFL outperforms existing models such as KNN with 1.75%, against Decision Tree with 3.44%, against Gaussian Naïve Bayes with 1.85%, against Logistic Regression with 3.63%, against Random Forest with 0.52%, against FNN with 6.32% and against SVM with 3.13%. With Correlation Model FNN-SFL outperforms existing models such as KNN with 0.37%, against Decision Tree with 0.68%, against Gaussian Naïve Bayes with 2.89%, against Logistic Regression with 4.80%, against Random Forest with 3.21%, against FNN with 6.42% and against SVM with 3.11%.

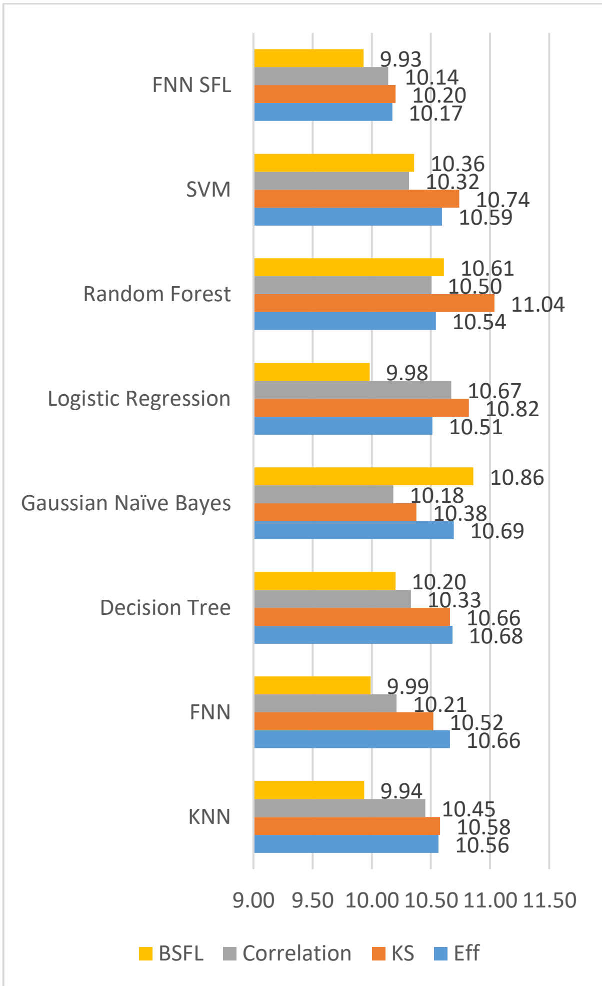

Figure 12. Comparison graph of FNN-SFL with existing methods w.r.t. Miss ratio on classification of grooming documents

With BSFL, FNN-SFL outperforms existing models such as KNN with 1.96%, against Decision Tree with 0.22%, against Gaussian Naïve Bayes with 3.5%, against Logistic Regression with 0.08%, against Random Forest with 4.96% and against SVM with 2.88%.

In Table 12 the comparison model of FNN-SFL with existing methods w.r.t. miss ratio on classification of grooming Documents has been given. This table shows the significance of FNN-SFL on miss ratio over other existing models

In Detail of Figure 12 with respect to Specificity on classification FNN-SFL outperforms the existing models along with the feature selection model Eff. The significance in terms of percentage against KNN with 3.70%, against Decision Tree with 4.75%, against Gaussian Naïve Bayes with 4.86%, against Logistic Regression with 3.22%, against Random Forest with 3.48%, against FNN with 6.93% and against SVM with 3.97%. With KS Model FNN-SFL outperforms existing models such as KNN with 3.56%, against Decision Tree with 4.31%, against Gaussian Naïve Bayes with 1.71%, against Logistic Regression with 5.72%, against Random Forest with 7.57%, against FNN with 8.21% and against SVM with 5.01%. With Correlation Model FNN-SFL outperforms existing models such as KNN with 2.99%, against Decision Tree with 1.84%, against Gaussian Naïve Bayes with 0.42%, against Logistic Regression with 4.97%, against Random Forest with 3.46%, against FNN with 6.36% and against SVM with 1.70%. With BSFL, FNN-SFL outperforms existing models such as KNN with 0.05%, against Decision Tree with 2.64%, against Gaussian Naïve Bayes with 8.55%, against Logistic Regression with 0.53%, against Random Forest with 6.40% and against SVM with .4.13%.

In Detail with respect to accuracy on classification FNN-SFL outperforms the existing models along with the feature selection model Eff. The significance in terms of percentage against KNN with 4.89%, against Decision Tree with 0.24%, against Gaussian Naïve Bayes with 3.5%, against Logistic Regression with 5.42%, against Random Forest with 3.69%, against FNN with 0.73% and against SVM with 0.04%. With KS Model FNN-SFL outperforms existing models such as KNN with 0.19%, against Decision Tree with 4.97%, against Gaussian Naïve Bayes with 2.52%, against Logistic Regression with 9.09%, against Random Forest with 3.94%, against FNN with 0.93% and against SVM with 4.45%. With Correlation Model FNN-SFL outperforms existing models such as KNN with 2.19%, against Decision Tree with 3.74%, against Gaussian Naïve Bayes with 6.36%, against Logistic Regression with 9.48%, against Random Forest with 5.12%, against FNN with 1.96% and against SVM with 7.49%.

5. CONCLUSION

In this paper a detailed analysis on the working model of classification with the fusion of FNN and SFL has been done. The results show the significance of the proposed model in terms of accuracy, precision, recall, sensitivity, specificity and miss ratio. With BSFL, FNN-SFL outperforms existing models such as KNN against Decision Tree, Gaussian Naïve Bayes, Logistic Regression, Random Forest and against SVM. Hence the research gives the clear idea on classification of judgement cases in which the future model can be extended with phenomenal accuracy over the classification model on judgement case files.

REFERENCES

[1] Mani, I., Maybury, M.T. (1999). Advances in Automatic Text Summarization. Cambridge MIT Press.

[2] Domingos, P., Pazzani, M. (1997). On the optimality of the simple Bayesian classifier under zero-one loss. Machine Learning, 29: 103-130. https://doi.org/10.1023/A:1007413511361

[3] Breslow, L.A., Aha, D.W. (1997). Simplifying decision trees: A survey. Knowledge Engineering Review, 12(1): 1-40. https://doi.org/10.1017/S0269888997000015

[4] Zaki, M.J., Jr Meira, W. (2014). Data Mining and Analysis: Fundamental Concepts and Algorithms. Cambridge University Press.

[5] Cover, T., Hart, P. (1967). Nearest neighbor pattern classification. IEEE Transactions on Information Theory, 13(1): 21-27. https://doi.org/10.1109/TIT.1967.1053964

[6] Wettschereck, D., Aha, D.W., Mohri, T. (1997). A review and empirical evaluation of feature weighting methods for a class of lazy learning algorithms. Artificial Intelligence Review, 11: 273-314. https://doi.org/10.1023/A:1006593614256

[7] Yang, Y., Webb, G.I. (2003). On why discretization works for Naive-Bayes classifiers. Lecture Notes in Computer Science, 2903: 440-452. https://doi.org/10.1007/978-3-540-24581-0_37

[8] Okamoto, S., Yugami, N. (2003). Effects of domain characteristics on instance-based learning algorithms. Theoretical Computer Science, 298(1): 207-233. https://doi.org/10.1016/S0304-3975(02)00424-3

[9] Kubat, M., Cooperson, M. (2001). A reduction technique for nearest-neighbor classification: Small groups of examples. Intelligent Data Analysis, 5(6): 463-476. https://doi.org/10.3233/ida-2001-5603

[10] Brighton, H., Mellish, C. (2002). Advances in instance selection for instance-based learning algorithms. Data Mining and Knowledge Discovery, 6: 153-172. https://doi.org/10.1023/A:1014043630878

[11] Wilson, D.R., Martinez, T. (2000). Reduction techniques for instance-based learning algorithms. Machine Learning, 38: 257-286. https://doi.org/10.1023/A:1007626913721

[12] Jensen, F. (1998). An introduction to Bayesian networks. Journal of the Royal Statistical Society. Series D (The Statistician), 47(2): 397-398.

[13] Bouckaert, R.R. (2004). Naive Bayes classifiers that perform well with continuous variables. Lecture Notes in Computer Science, 3339: 1089-1094. https://doi.org/10.1007/978-3-540-30549-1_106

[14] Vishwanathan, S.V.M., Murty, M.N. (2002). SSVM: A simple SVM algorithm. Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No.02CH37290), Honolulu, HI, USA, pp. 2393-2398. https://doi.org/10.1109/IJCNN.2002.1007516

[15] Soucy, P., Mineau, G.W. (2001). A simple KNN algorithm for text categorization. Proceedings 2001 IEEE International Conference on Data Mining, San Jose, CA, USA, pp. 647-648. https://doi.org/10.1109/ICDM.2001.989592

[16] Sugumaran, V., Muralidharan, V., Ramachandran, K.I. (2007). Feature selection using decision tree and classification through proximal support vector machine for fault diagnostics of roller bearing. Mechanical

Systems and Signal Processing, 21(2): 930-942. https://doi.org/10.1016/j.ymssp.2006.05.004
[17] McGeachie, M.J., Chang, H.H., Weiss, S.T. (2014). CGBayesNets: Conditional Gaussian Bayesian network learning and inference with mixed discrete and continuous data. PLoS Computational Biology, 10(6): e1003676. https://doi.org/10.1371/journal.pcbi.1003676
[18] Hosmer, D.W., Lemeshow, S. (2013). Applied Logistic Regression. John Wiley & Sons. https://doi.org/10.1002/0471722146
[19] Ho, T.K. (1995). Random decision forests. Proceedings of 3rd International Conference on Document Analysis and Recognition, Montreal, Quebec, Canada, pp. 278-282. https://doi.org/10.1109/ICDAR.1995.598994