

# Enhancement in Performance of Financial Crisis Prediction using Hybridization of Machine Learning Classifiers

S. Anand Christy, R.ArunKumar

**Abstract:** Financial Crisis has been the stern problem experienced by various organizations or even common people when interested in investing in any Financial institutions like banks, Funds development institutions etc. Hence it is mandatory that a reliable prediction system should be applied in early prediction of Financial Crisis Prediction thereby preventing investment in weak financial institutions that might lead to bankruptcy. The Paper focuses on designing a Hybrid Optimized Algorithm called Hybrid Unified Machine Classifier (HUMC) based on Machine Learning Technique that would be capable of identifying categorized and continuous variables in a financial crisis dataset and determine the confusion matrix that can be instilled in performance analysis tool comprising of analytics and prediction related to Accuracy, F-Score, Sensitivity, Specificity, False Positive Rate (FPR) and False Negative Rate (FNR) respectively. Early testing with the training set of Australian credit dataset were tested with machine learning classifiers like Decision Tree, PART, Naive Bayesian, RBF Network and Multilayer Perceptron algorithms with accuracies 85.50%, 83.62%, 77.24%, 82.75% and 84.93% respectively. The Algorithm HUMC was developed based on combining classification features from decision tree, identifying hidden nodes and model with boosting technique that could enhance the performance levels of the Financial Crisis Prediction. The design of algorithm comprised of best characteristics of both classification and neural networks that are capable to find categorization criteria in the dataset at the first level and also to find the hidden continuous data during the second stage respectively. The design of HUMC was implemented and tested with MATLAB. The Result showed that HUMC algorithm showed greater accuracy (86.25%) in comparison to other classifier models along with other performance measures. Thus, this algorithm enhances the prediction of Financial Crisis predictions with good performance.

**Keywords:** Hybrid Unified Machine Classifier, Machine Learning Algorithms, Financial Crisis Prediction, classification algorithms, performance enhancement.

## I. INTRODUCTION

Financial Crisis Prediction (FCP) is one of the major areas where high level organization to low level investors are keen to understand and predict the results with enhanced accuracy. The impact of investing on the Financial institutions that might prolong for a longer period with high assets and cash

Revised Manuscript Received on January 5, 2020

\* Correspondence Author

**S. Anand Christy\***, Department of Computer and Information Science, Annamalai University, Chidambaram, Cuddalore, TamilNadu, India.  
Email: scholarchristy83@gmail.com

**Dr.R.Arunkumar**, Department of Computer Science and Engineering, Annamalai University, Chidambaram, Cuddalore, TamilNadu, India.  
Email: arunkumar\_an@yahoo.com

flows in differentiation with the organizations that were on the verge of getting bankrupted in course of time. The Financial Crisis prediction is based on the assets and liabilities of the organization that has to be assessed with right threshold values to predict the accuracies. The major objective is to design a Hybrid Optimized Algorithm based on Machine Learning Technique and optimization algorithms that are capable to predict the Financial Crisis of an organization. This method subjects to identify categorized and continuous variables in a financial crisis dataset and determine the confusion matrix. The importance of enhancement in performance was attributed to reliability of the prediction from potential investors as well.

The Problem to be addressed has been the early prediction of Financial crisis of an organization through combination of machine learners, optimization methods and classifier models. Similar Financial Crisis predictions were analyzed by (Hong Hanh Le & Jean-Laurent Viviani, 2018) through comparison of accuracies and performances of statistical techniques and machine learning techniques. A comparison on 3000 banks in US was experimented with Discriminant Analysis and Logistic Regression from statistical methods and ANN, SVM, K-Nearest Neighbor in-terms of machine learning approaches to identify the best features that determine the performance of prediction of financial crisis data. It is identified that ANN and K-nearest neighbor model from machine learning methods were more accurate comparing to other statistical models. Thus, the features of Financial Crisis play a significant role in predicting the future occurrence of crisis at the earlier stage itself.

## II. RELATED WORKS

Financial Crisis prediction had been conducted earlier with machine learning classifiers using various hybrid models. Many researches work focus on Financial crisis predictions captivated through machine learning techniques and enhancement is recommended in few cases. (Yu-Pei Huang and Meng-Feng Yen, 2019) conducted a review on all the machine learning techniques that are useful in predicting Financial Crisis of an organization. Among the researched supervised, unsupervised and hybrid models, four models including Support Vector Machines (SVM), Hybrid Associative Memory with Translation, Hybrid GA-fuzzy cluster model and extreme Gradient Boosting (XGBoost) algorithms found to have high accuracy.

# Enhancement in Performance of Financial Crisis Prediction using Hybridization of Machine Learning Classifiers

Thus, Financial Crisis predictions were possible using machine learning methods. The existing problem also focuses on the prediction of Financial Crisis with a comparative analysis of traditional machine learning models and a novel hybrid model. A Model based on predicting the crisis in currency values was also proposed using Deep Neural decision tree methods (David Alaminos et.al, 2019).

(Arian Dhini et.al 2019) combined the features of clustering algorithms like k-means with classification techniques to develop a hybrid model that was capable of predicting the existence of Financial Crisis in organisations. The results are indicated as two clusters with 17% in one cluster and 83% in another. Further, the outcomes are tested based on clusters and obtained an accuracy of 98.611% with logistic regression. Another result obtained with C4.5 decision Tree along with boosting achieved an AUC value of 0.996 and 0.990 output. Thus, combination of algorithms could enhance the prediction accuracy of complex manipulations like Financial Crisis predictions.

A more complicated analysis with machine learning classifiers was performed by (Ching-Hsue Cheng, Chia-Pang Chan, and Jun-He Yang, 2018) using seasonal time-series gene expression programming model as traditional models were devoid of time-series methods and feature selection models were neglected. The novel time series model applied mathematical rules and attribute selections to enhance decisions in financial crisis. (Ziyan Yuan and Yan Hou, 2019) designed an Intelligent Decision-making technique that was capable of predicting early warning signs in Financial crisis through theoretical analysis.

(Deron Liang et.al, 2018) generated a classifier ensemble model that was capable of reducing the errors that occurred during prediction using unanimous voting (UV) technique. (Thi Kha Nguyen and Thi Phuong Trang Pham, 2019) designed an ensemble model based on Artificial Intelligence to predict financial crisis of an organisation based on two benchmark datasets. The outcomes showed that bagging models produced a higher accuracy comparing to other models. These research works suggested that hybrid models that are formed as combination of machine learning models could outperform many traditional models as they lack in some of the criterion to predict the financial crisis. This verdict was the backbone of seal the gap of this research study.

### III. PREDICTION WITH MACHINE LEARNING CLASSIFIERS

Machine Learning is the ability of the system to learn the rules based on the input and output content. Machine learning combined with optimization techniques were capable of improving the performance. A hybrid combination of machine learning techniques with optimization was performed by (J. Uthayakumar et.al 2018) which is the basis to kindle a hope for hybrid optimisation algorithm called Fitness-Scaling Chaotic Genetic Ant Colony Algorithm (FSCGACA) and used it with k-means clustering to test three benchmark datasets in financial crisis. The research completed tests with machine learners like logistic regression, RBF Network and Multilayer Perceptron to achieve an

accuracy of 99.20. Further testing with Ant Miner and rule set generation was also successful. However, the classification of categorical data was not possible in the approach. This research problem is the motivation for the current research.

The initial benchmark dataset used for the prediction called Australian Credit was pre-processed to remove outliers and non-numeric data. The Australian Credit was identified as one of the complicated datasets as revealed by () to predict financial crisis as it has unlabeled and hidden data. The data formats were a combination of categorical as well as continuous data. The categorical data denoted the data that are capable to be represented as classified or data that can be categorized based on some criteria. The continuous data can be within a range of values that deviated from one another. Hence handling such data required good algorithms like decision tree algorithm, naive Bayesian algorithm, PART algorithm, RBF Network and Multi-Layer Perceptron to test in initial cases. The dataset was loaded and with 5 cross folds the results are identified as tabulated in Table 1.

**Table-1: Performance measures of various Traditional Machine Learning classifiers**

Classifier	MLP	RBF Network	Naive Bayesian	PART Decision	Decision Tree
Accuracy	85.07	85.36	79.13	83.88	84.92
FPR	0.156	0.136	0.095	0.182	0.156
FNR	0.143	0.153	0.258	0.95	0.145
Sensitivity	0.856	0.846	0.741	0.0483	0.854
Specificity	0.843	0.863	0.904	0.817	0.843
F-score	86.42	86.93	83.17	9.15	86.31
Kappa	0.698	0.703	0.567	0.673	0.695

The Results signify that the Australian Credit though complicated could be predicted using traditional machine learning models. However, it was required to obtain a refinement of performances in all these algorithms through optimization methods. Hence a hypothesis is proposed that a hybrid method developed as a combination of all these models would be capable of enhancing the accuracy and other performance measures like False Positive Rate, False Negative Rate, Sensitivity, Specificity, F-Score and Kappa values in a significant manner.

### IV. DESIGN OF HYBRID UNIFIED MACHINE CLASSIFIER

The Traditional models showed a good performance with accuracies. However, there was a need to enhance the accuracy levels of the financial crisis prediction using hybridization of algorithms. Earlier, a forecasting model similar to the current model was developed by (Mohammed Siddique, Sabyasachi Mohanty, Debdulal Panda, 2018) to propose a combination machine learning model like Support Vector Regression and optimization model like Ant Colony Optimization (ACO) respectively. The model was built and tested with dataset where the results showed optimal performance.

A Similar framework has been found in (Tingyu Zhang et.al, 2018) where logistic regression, support vector machine and RBF network were combining to predict landslide susceptibility. The Hybrid model of logistic regression with decision tree was designed by (Arno De Caignya, Kristof Coussementa, Koen W.De Bock, 2018) to detect customer churn in financial organisations. Hence a hybrid of statistical methods like Multilayer Perceptron, RBF network with machine learning models like decision tree, naïve Bayesian and PART algorithm was possible.

A Novel Machine Learning model, Hybrid Unified Machine Classifier (HUMC) is proposed as a progression of the existing hybrid models to predict the Financial Crisis based on hidden and classification patterns of data in the Australian Credit dataset. This model combines the best features of Classification based machine learners like decision tree algorithm and hidden node identifying neural and statistical model like Multilayer Perceptron to enhance the accuracy of prediction.

#### 4.1 Combined Models of HUMC Algorithm

The HUMC Model combines both machine learning and statistical neural models with optimization of fitness functions to determine the best outcome in terms of accuracy and other performance measures. The combination has been based on the stages that depicts utilization of best characteristics of each of the model and enhance it for prediction. The decision tree was a classification algorithm capable of categorizing data based on the input values. It also refines the data using entropy and other boosting techniques. Decision tree are capable of generating rulesets, handle categorical data just by matching without much computations. Though they can handle continuous data, they are error prone even after pre-processing which leads to a deficit in predicting Financial Crisis. However, the pitfalls in decision tree can be managed using Multilayer perceptron which comprised of three layers to identify hidden nodes missed in decision trees. The neuron in each node used a nonlinear function like least mean square method that would train the input and output to predict the anonymous data in the dataset as indicated in Eq. (1).

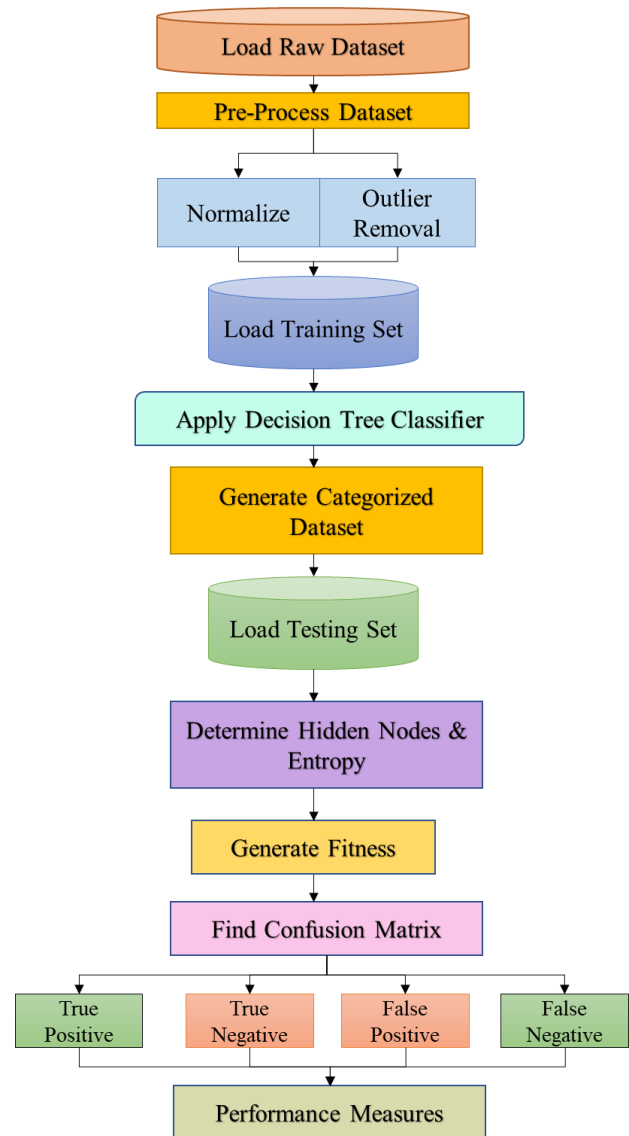
$$net(i) = w_{i0} + \sum_{j \in pred(i)}^{\infty} (w_{ij} a_j) \text{-----} (1)$$

Where net(i) is the network input, a is the activation output, w refers to weights of nodes and pred(i) represents predecessors of hidden nodes respectively. This feature of multilayer perceptron has been utilized with decision tree classification, entropy and threshold detection to form a hybrid algorithm called Hybrid Unified Machine Learner (HUMC) Model. Thus, combination of best features of two models provide solution for predicting complex datasets like that of financial crisis predictions.

#### 4.2 Structure of HUMC Model

Hybrid Unified Machine Learner (HUMC) is modeled as a combination of classification of categorical features using decision tree algorithm and identification of hidden features and data using Multilayer perceptron that is capable of learning from unlabeled data sources. The overall architecture

of HUMC comprised of 5 stages as shown in Figure 1. The initial stage loads the raw dataset for preprocessing where normalization techniques and outlier removal methods are applied to removed non-numeric and unrelated data in the dataset. Being the preliminary stage, it contains unlabeled and correlated data. During Second Stage, the data has been applied with decision tree models to categorize the data in the dataset and form a testing set. The testing set comprised of ordered data of 14 attributes whose identity is unknown.



**Figure 1: Architecture of Hybrid Unified Machine Classifier (HUMC) Model**

Hence during the third stage, the testing set is applied with feature combinations of multilayer perceptron to measure the weights and determine the threshold values of individual categorized features. The manipulated data is computed for entropy for applying the results to fitness function. The fourth stage retrieves the outputs from fitness function and determines the four parameters, True Positive, True Negative, False Positive and False negative respectively.



The Final Stage intakes the confusion matrix values to compute the performance measures used in the study viz Accuracy, False Positive Rate (FPR), False Negative Rate (FNR), Sensitivity, Specificity, F-score and Kappa values respectively. The outcomes received has been compared with the existing machine learning models to determine its performance and efficiency. The Algorithm for HUMC is given in Table 2.

**Table 2: Algorithm for Hybrid Unified Machine Learner (HUMC) Financial Crisis Prediction**

```

Hybrid Unified Machine Learner (HUMC)
Algorithm
Input: Australian Credit raw Dataset A (), A1(), A2()
1: Define  $a \leftarrow \text{Attributes}, c \leftarrow \text{Class}$ 
2: Initialize  $n, i \leftarrow 1, j \leftarrow 1, tp \leftarrow 1, tn \leftarrow 1, fp \leftarrow 1, fn \leftarrow 1$ 
3:  $s \leftarrow \text{size}(\text{Attributes}), t \leftarrow \text{Threshold}$ 
//Pre-Processing
4: For  $\forall i \in n$  do
5: For  $\forall j \in s$  do
6: Normalize (i,j)
7: Outlier_detect (i,j)
8: End For
9: End For

// Data Categorization
10: For  $\forall i \in n$  do
11: For  $\forall j \in s$  do
12: If  $f(i,j) < t$ 
13: Group A1()
14: Else
15: Group A2()
16: End if
17: End For
18: End For

// Hidden Data computation
19: For  $\forall i \in n$  do
20: For  $\forall j \in s$  do
21: If (A1(i, j) matches A2(i, j)) then
22: determine new A()
23: End If
24: End For
25: End For
26: End for

//Fitness Calculation
27: For  $\forall i \in n$  do
28: Test A(i)with t
29: If A1()  $\leq t$ 
30: tp++;
31: Else
32: tn++;
33: End if
34: If A1()  $\leq t$ 
35: fp++;

```

```

36: Else
37: fn++;
38: End if
39: End for
40: Calculate Accuracy, FScore, Kappa

```

The algorithm utilized the initial parameters A () which is the raw dataset was subjected to preprocessing techniques like outlier removal and normalization process. The preprocessed dataset is categories into two classified datasets A1() and A2() based on the threshold (t) value to determine the assets and liabilities categories of the dataset in financial crisis. The dataset is further compounded to hidden nodes test to identify the unknown or unlabeled data available in the dataset. The data had been obtained and presented in the new dataset which is further tested for entropy and fitness calculation. The final outcome was classified into four categories of True Positive (Bankrupt as Bankrupt), True Negative (Bankrupt as Non-Bankrupt), False Positive (Non-Bankrupt as Bankrupt) and False Negative (Non-Bankrupt as Non-Bankrupt) respectively. The classified data was formed as confusion matrix which was further undergone to performance test using a tool designed to find accuracy, FPR, FNR, F-Score and Kappa values to have a comparative analysis with traditional machine learning models.

### V. IMPLEMENTATION OF HUMC CLASSIFIER MODEL

The Design of HUMC classifier model was implemented in MATLAB 2015 and the initial dataset is loaded. The loaded dataset comprised of 14 features of Australian Credit dataset with no further refinement. The dataset is refined and pre-processed as indicated in Figure 2. The preprocessing indicated the impurities in the dataset that had to be removed and also the format to be changed during the process. The modified dataset formed the training set ready for categorization process.

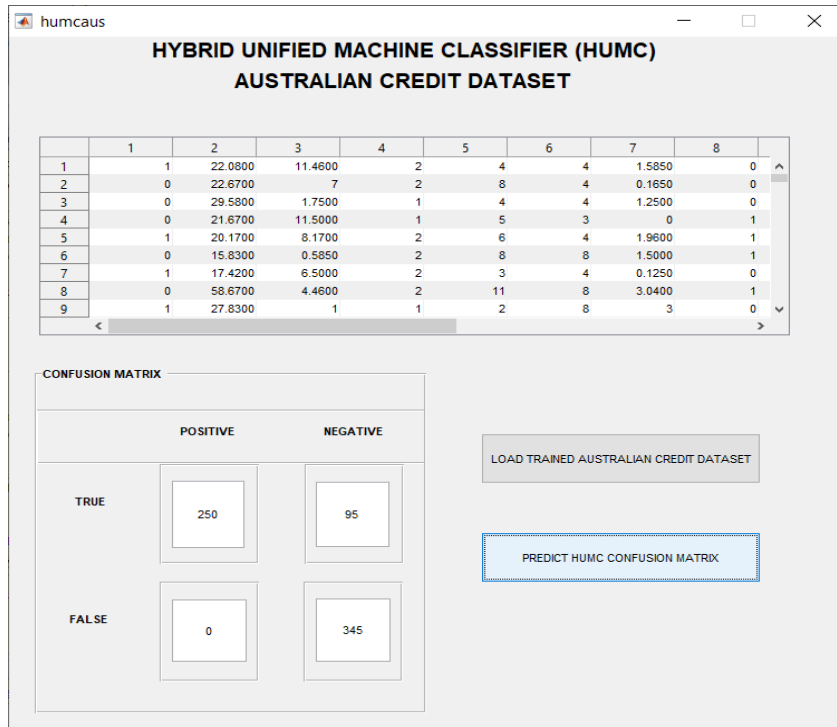
The preprocessed dataset was loaded in the interface and applied with HUMC algorithm. The dataset was initially categories based on the ruleset and threshold values using decision tree and then identified with unlabeled data based on multilayer perceptron values respectively. The result was obtained on screen based on the dataset in the form of a confusion matrix as depicted in Figure 3. After HUMC Outcomes as confusion matrix, the further testing is concluded based on performance analysis that assists in conducting analytics with traditional machine learning benchmark algorithms used in the study, Decision Tree, Naive Bayesian, PART algorithm, RBF Network and Multilayer Perceptron. The results of confusion matrix are further loaded into performance analysis tool and determined with performance measures as indicated in Figure 4.

The outcomes of accuracy and other measures of HUMC are presented in Table 2.

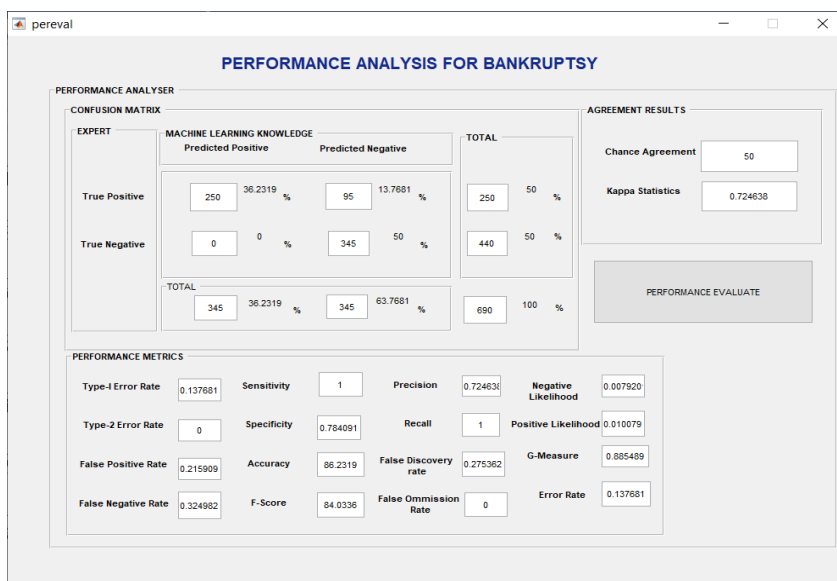
**Table 2: Performance measures of HUMC model**

Classifier	HUMC
Accuracy	86.23
FPR	0.215
FNR	0.325
Sensitivity	1
Specificity	0.7841
F-score	84.0336
Kappa	0.724

It has been obligatory that the proposed model has to be tested with other traditional models selected in the research study to indicate the best model in terms of performance and efficiency.



**Figure 3: Confusion Matrix outcome of HUMC after execution**



**Figure 4: Performance Analysis of HUMC based on Confusion Matrix input**

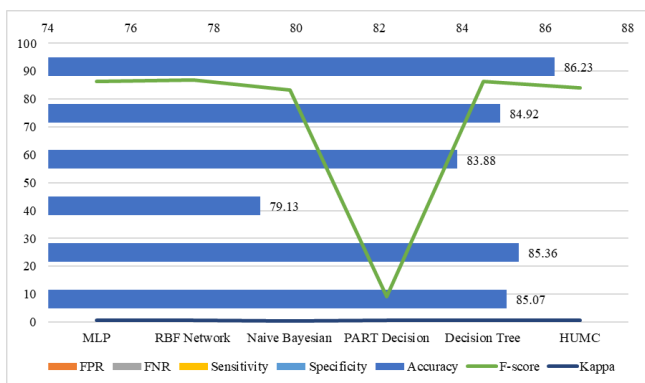
## VI. RESULTS AND DISCUSSION

The Proposed HUMC model has been compared with the performance levels of other machine learning algorithms to provide support to the hypothesis that hybrid models provide enhanced performance in comparison to individual algorithms for a specific purpose. The comparative analysis of algorithms tested with Australian Credit dataset and their

performance analysis are provided in Table 3. The table indicated a steady increase in the accuracy levels of the prediction accuracy in Hybrid Unified Machine Classifier model with 86.23 in comparison to other models MLP (85.07), RBF Network (85.36), Naive Bayesian (79.13), PART Decision (83.88), Decision Tree (84.92) as indicated in the Figure 5.

**Table 3: Comparative analysis of machine learning models with proposed hybrid model**

Classifier	MLP	RBF Network	Naive Bayesian	PART Decision	Decision Tree	HUMC
<b>Accuracy</b>	85.07	85.36	79.13	83.88	84.92	86.23
<b>FPR</b>	0.156	0.136	0.095	0.182	0.156	0.215
<b>FNR</b>	0.143	0.153	0.258	0.95	0.145	0.325
<b>Sensitivity</b>	0.856	0.846	0.741	0.0483	0.854	1
<b>Specificity</b>	0.843	0.863	0.904	0.817	0.843	0.7841
<b>F-score</b>	86.42	86.93	83.17	9.15	86.31	84.0336
<b>Kappa</b>	0.698	0.703	0.567	0.673	0.695	0.724



**Figure 5: Comparison of Hybrid and Traditional Models**

It is identified that apart from Accuracy, the sensitivity and specificity indicating the identification of bankrupt as bankrupt and non-bankrupt as non-bankrupt was also convincing. The F-Score based on precision and Recall also showed considerable improvement. The enhancement was noticed in kappa values with a maximum of 0.724. However, the model had an oscillation in detecting False positive rate and False negative rate but in comparison to other models, the results are convincing. Considering the above performance and efficiency levels, the proposed hybrid model was found to be reliable in comparison to traditional models and the results are obtained to contemplate to the Prediction methods in Financial Crisis of an organization.

## VII. CONCLUSION

This study has developed a hybrid model for predicting Financial Crisis of an organization that was objected to combine the best features of the classification based machine learning algorithms like decision tree algorithm to handle categorical data and statistical based machine learners like Multilayer perceptron to handle hidden data and also to handle continuous data which is not possible in normal decision tree algorithms to enhance the performance levels of the predictions. The proposed hybrid model HUMC

outperformed other traditional models built for a specific purpose in terms of accuracy, rule generation as well as factors like sensitivity, specificity, F-Score and kappa respectively. The hybrid model affirms a reliable method that could be applied with future datasets of financial organizations to determine the prediction of bankruptcy at an earlier stage so that investors and economy of common people could be protected. This algorithm has scope to be implemented in bioinspired models and expert systems for predictions in future.

## REFERENCES

- G. O. Dhini, A., Aji, N. A., Putri, H. R., & Indriyanti, D. (2019, July). Hybrid Classifier for Predicting Financial Distress. In 2019 16th International Conference on Service Systems and Service Management (ICSSSM) (pp. 1-5). IEEE.
- Alaminos, D., Becerra-Vicario, R., Fernández-Gámez, M. Á., & Cisneros Ruiz, A. J. (2019). Currency Crises Prediction Using Deep Neural Decision Trees. *Applied Sciences*, 9(23), 5227.
- Liang, D., Tsai, C. F., Dai, A. J., & Eberle, W. (2018). A novel classifier ensemble approach for financial distress prediction. *Knowledge and Information Systems*, 54(2), 437-462.
- Nguyen, T. K., & Pham, T. P. T. (2018). Predicting bankruptcy using machine learning algorithms.
- Yuan, Z., & Hou, Y. (2019, June). Research on Intelligent Decision-Making Method of Enterprise Financial Crisis Early Warning. In *Journal of Physics: Conference Series* (Vol. 1213, No. 2, p. 022026). IOP Publishing.
- Cheng, C. H., Chan, C. P., & Yang, J. H. (2018). A Seasonal Time-Series Model Based on Gene Expression Programming for Predicting Financial Distress. *Computational intelligence and neuroscience*, 2018.
- Uthayakumar, J., Metawa, N., Shankar, K., & Lakshmanprabu, S. K. (2018). Financial crisis prediction model using ant colony optimization. *International Journal of Information Management*.
- Huang, Y. P., & Yen, M. F. (2019). A new perspective of performance comparison among machine learning algorithms for financial distress prediction. *Applied Soft Computing*, 83, 105663.
- Chatzis, S. P., Siakoulis, V., Petropoulos, A., Stavroulakis, E., & Vlachogiannakis, N. (2018). Forecasting stock market crisis events using deep and statistical machine learning techniques. *Expert systems with applications*, 112, 353-371.

10. Uthayakumar, J., Metawa, N., Shankar, K., & Lakshmanaprabu, S. K. (2018). Intelligent hybrid model for financial crisis prediction using machine learning techniques. *Information Systems and e-Business Management*, 1-29.
11. Le, H. H., & Viviani, J. L. (2018). Predicting bank failure: An improvement by implementing a machine-learning approach to classical financial ratios. *Research in International Business and Finance*, 44, 16-25.
12. Siddique, M., Mohanty, S., & Panda, D. (2018). A hybrid forecasting model for prediction of stock value of tata steel using support vector regression and particle swarm optimization. *International Journal of Pure and Applied Mathematics*, 119(14), 1719-1727.