IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 29 August 2022, accepted 19 September 2022, date of publication 26 September 2022, date of current version 10 November 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3210032

RESEARCH ARTICLE

Predicting Insolvency of Insurance Companies in Egyptian Market Using Bagging and Boosting Ensemble Techniques

AHMED A. KHALIL[1,2], ZAIMING LIU[1], AHMAD SALAH[3,4], AHMED FATHALLA[5], AND AHMED ALI[6,7]

[1]School of Mathematics and Statistics, Central South University, Changsha, Hunan 410017, China
[2]Faculty of Commerce, Assiut University, Asyut 71515, Egypt
[3]Department of Computer Science, Faculty of Computers and Informatics, Zagazig University, Zagazig 44519, Egypt
[4]College of Applied Science, University of Technology and Applied Sciences, Ibri 115, Oman
[5]Department of Mathematics, Faculty of Science, Suez Canal University, Ismailia 41522, Egypt
[6]Department of Computer Science, College of Computer Engineering and Sciences, Prince Sattam Bin Abdulaziz University, Al-Kharj 16278, Saudi Arabia
[7]Higher Future Institute for Specialized Technological Studies, Cairo 3044, Egypt

Corresponding authors: Ahmad Salah (ahmad@zu.edu.eg) and Ahmed Fathalla (fathalla_sci@science.suez.edu.eg)

ABSTRACT Insolvency is a crucial problem for several insurance companies that suffer from it. This problem has direct or indirect effects on both the people working in the financial business and normal citizens. Thus, in insurance companies, the ability to predict insolvency is in great demand. There are several efforts proposed to predict insurance company insolvency using computer science methods (e.g., support vector machine and fuzzy systems). Each country has its own data patterns due to interior matters. Thus, insurance companies from different countries may have different data patterns. Consequently, the utilized predictive model should adapt to the dataset at hand. To our best knowledge, despite there are several efforts to build an insolvency predictive model, none of these efforts explored the Egyptian market. In addition, even the existing efforts did not utilize the ensemble learning methods in the insolvency prediction problem. In this context, we have two main contributions to this work. First, we proposed the first public access dataset of Egyptian insurance companies. The collected dataset was gathered from 11 Egyptian insurance companies during the years 1999 to 2019. The dataset consists of a set of 22 ratios (21 input features and one output feature), e.g., retention and investment yield alongside the solvency ration (i.e., the target feature). In the second contribution, we proposed exploring the performance of the ensemble learning methods to address the insolvency prediction problem. Thus, we proposed building several insolvency predictive models using ensemble learning and classic machine learning models. Next, the proposed models are evaluated on different accuracy metrics, e.g., Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). The experimental results revealed that the ensemble learning-based models outperformed the classic machine learning-based models. Moreover, the correlation analysis between the utilized 22 financial ratios revealed that the most significant ratios, for the task of predicting the solvency ratio, are the technical provisions to shareholders' funds, insurance companies' debit balances to shareholders, and earnings after taxes to shareholders' funds.

INDEX TERMS Bagging, Egyptian market, ensemble models, insolvency, insurance, machine learning.

I. INTRODUCTION

In any society, the insurance sector plays a significant role in maintaining economic stability as a major financial market investor. Besides, insurance companies play a significant role in offering essential financial services for society's members. These companies insure the insured against various Financial threats that the insured may face through insurance coverage that allows households, corporations, and public sector entities to transfer these risks to the insurance companies.

The associate editor coordinating the review of this manuscript and approving it for publication was Sabah Mohammed.

For example, general insurance companies assist companies and households in limiting the financial costs of different risks to their personal belongings, legal liability, etc. In some instances, insurance companies on both sides of the balance sheet face various threats, and because of deteriorating financial market conditions, the value of an insurance company's assets can decrease.

Meanwhile, distress or the failure of an insurance company may occur either because of insufficient claims provisions or inadequate capital to resist unanticipated losses, or because of insured incidents or fluctuations in assets it holds. Some insurance companies' activities may increase the risk of trouble or loss and ultimately cannot regain their position, which is referred to as the problems of financially distressed or financial insolvency [1]. Consequently, potential insolvencies need to be identified as soon as possible to direct regulators in the insurance companies towards deficiencies that require attention to prevent actual insolvencies [2], [3].

Insolvency is one of the critical threats that insurance companies aim to eliminate because of its negative impacts, which may affect the insurance industry's continuity. When an insurance company becomes insolvent, it will influence the workers and put the shareholders' and policyholders' stakes at risk. As a result, insolvency has become a significant concern for insurance companies. Thus, it is critical for them that early warning signs of financial distress can be identified in a company before insolvency [4], [5].

The insurance insolvency phenomenon is global, and so many approaches for predicting insolvency in the insurance industry have been suggested. The prior research, which focused on financial insolvency in property-liability insurers, is divided into two directions. First, some research focuses exclusively on evaluating and forecasting insolvency using financial indicators as variables, but such variables generally violate statistical assumptions [6], [7], [8], [9]. Second, some studies proposed statistical methods, such as Canonical correlation analysis, discriminant or logistic analysis [10], [11], [12], and other studies tried to use Artificial Intelligent (AI) based methods such as neural networks, fuzzy kernel c-means, and rough set theory [4], [13], [14], [15], [16], [17], [18], [19], [20], [21]. Recently, some studies attempted to utilize machine learning models to predict insurance insolvencies, such as support vector machine (SVM) and random forest (RF), and compared them with statistical models; the findings showed higher accuracy in machine learning models [3], [22]. Despite these efforts, little research has been conducted to predict insurance insolvencies in the Egyptian market. In addition, there is no research utilizing the ensemble learning technique.

Our research adds in several ways to the existing literature on insurance insolvency. First, defining the main and important ratios that indicate the insolvency of insurance companies in the Egyptian insurance market can help decision-makers diagnose a company's likelihood of survival and avoid that risk. Second, in order to enhance the predictability and to help build an early warning system capable of being used by regulators for insolvency prediction and priority ranking of insurance companies, in this research, we propose bagging and boosting ensemble techniques to predict the insurance companies' insolvency based on a real dataset of Egyptian insurance companies in the Egyptian market. There has been a lack of empirical studies about this matter in the Egyptian market. Besides, our proposed work utilizes the bagging and boosting methods for the first time in solving the problem of predicting the insolvency of insurance companies, as the existing literature studies utilized machine learning techniques such as random forest and support vector regressor (SVR). Thus, the main contributions of this work can be summarized as follows:

1) To the best of our knowledge, we contributed to the literature the first dataset of the Egyptian insurance market. This dataset can be utilized to compare the patterns of insolvency insurance companies from different countries. This help to link countries with similar insolvency events. The proposed dataset is made publicly available.[1]
2) We proposed studying the correlation of the different financial factors (e.g., collective ratio and operation ratio) to figure out the most significant factor(s) in the task of insolvency prediction of insurance companies.
3) We proposed exploring the performance gap between the classic machine learning (ML) methods and the ensemble learning methods in the task of predicting the insolvency of insurance companies on the proposed dataset.

The rest of this article is structured as follows. Background and literature review are presented in Section II, followed by data and methodology in Section III. Next, Section IV explains the experimental setup, evaluation metrics, and the findings of the experiment. Finally, the key findings and suggestions some directions for future research are offered in Section V.

II. BACKGROUND AND LITERATURE REVIEW

A. BACKGROUND

Machine learning and deep neural networks have recently exhibited top-notch performance in a wide range of applications [23], [24], [25], [26], For instance, various natural language processing (NLP) and computer vision tasks, such as language modeling [27], speech recognition [23], [24], computer vision [28], sentence classification [29], and machine translation [30], require the processing of text, images, and speech.

Moreover, machine learning has emerged as a novel way of financial data processing, which makes it an appropriate approach for predicting financial markets [31], [32], [33], [34]. From the machine learning perspective, the issue of insurance prediction can be formulated as a supervised machine learning task. In this context, machine learning

[1]https://github.com/Ahmed-Fathalla/Predicting-insolvency-of-insurance-companies-in-Egyptian

algorithms are powerful tools that can learn through experience and through the use of data without being explicitly programmed. Therefore, machine learning algorithms are widely used in different real-world applications, such as email filtering [35], sequence-to-sequence modeling [36], and computer vision [37].

Ensemble learning methods (bagging, boosting, and stacking) are important practical machine learning techniques used in different predictive problems [38], [39], [40], as they can consistently achieve higher predictive accuracy results compared to single powerful machine learning models. In ensemble learning techniques, multiple machine learning models (i.e., weak learners, or base learners) are trained to address the same predictive task, then they are integrated to gain the desired output.

In the Bagging and Boosting algorithms, the same learner-type algorithm is used. The major difference is that the training mechanisms of the weak learners in both methods are different. Bagging ensemble (stands for Bootstrap Aggregation) [41] is a type of ensemble learning technique that uses various datasets randomly sampled with replacements from the original dataset.

In such a way, the obtained datasets are used to train multiple base learners in parallel, and then, all of the outputs are averaged to determine the final result. A bagging ensemble is beneficial for reducing the variance and solving over-fitting issues in a model. Consequently, it improves the prediction ability of the weaker predictive models (regressor or classifier).

Boosting ensemble is a sequential method where the base-learners are trained sequentially and adaptively to impose more weight on the miss-classified (high error) observations to improve model predictions of a learning algorithm [42]. Boosting algorithms are beneficial for decreasing bias error and building strong predictive models.

In this paper, we propose predicting the insolvency of insurance companies in the Egyptian market using two ensemble methods, namely, 1) bagging and 2) boosting ensemble learning methods.

B. LITERATURE REVIEW

Insurance companies' insolvency is a critical and common occurrence that harms the economy, particularly in small to medium-sized enterprises, because of its adverse effects. We present in this section the previous studies that examine financial distress by developing models for forecasting insolvency based on key insurance company indicators in several ways, such as financial ratios (FR) as variables, statistical methods, and ML models.

Several previous research has sought to analyze and predict insolvency in property/liability insurance companies. Cummins [6], BarnNiv and Smith [7], BarNiv [8], and Harmelink [9] introduced business failure models based on financial ratios (FR) as explicative variables to evaluate and predict insolvency in the insurance industry. Where ratios may be used as financial analytical tools, they are not predictive since they normally depend on historical details. Again, while helping to systematically concentrate attention on key areas and to summarize knowledge in an understandable way and recognize patterns and relationships, such variables generally do not obey the assumptions of statistics. To circumvent these difficulties, some studies applied statistical methods such as canonical correlation analysis, logistic regression analysis, and Multiple Discriminant Analysis (MDA) to make predictions regarding the demise of insurance companies. For example, the MDA model correctly detects 92% of insolvent companies two years previous to the determination of their insolvency, according to Trieshmann and Pinches [12].

Ambrose and Carroll [10] compare and combine Best's ratings with financial ratios to determine their ability to predict life insurer insolvency by using MDA. Stowe and Watson [11] tried to reach a specific combination of assets that gives sufficient information about a specific combination of liabilities in order to examine the relationship between the structure of each of the assets and liabilities in life insurance companies by using Canonical correlation analysis.

Other studies on property-liability insurance financial insolvency tried to use Artificial Intelligent (AI) methods and machine learning models such as neural networks, fuzzy kernel c-means, random forest (RF), support vector machine (SVM), and rough set theory, for example. Brocket [13], [14], [16] tried to implement neural network methods to obtain an early warning of insurer insolvency or predict the insolvency of insurance companies, and the results of these neural network methods show high predictability and generalizability and outperform the traditional statistical approaches. Segovia Vargas *et al.* [15] applied the two methods (See5 and Rough Set) to the problem of forecasting the insolvency of Spanish non-life insurance companies using a set of financial ratios. The conclusion is that these methods can be an effective tool for determining an insurance firm's insolvency. Chiet *et al.* [18] created and developed an insolvency predictive model using ANN that could predict general insurance companies' future failure in Malaysia. The findings reveal a high degree of predictability. Rustam and Yaurita [22] tried to utilize SVM model and Rustam and Saragih [3] tried to use RF as a machine learning model to predict insurance companies' insolvency. According to the findings of two studies, machine learning models can be an effective tool for determining an insurance company's insolvency.

In [43], the authors addressed the problem of aggregated over-funded pension plans in the context of game theory. Thus, considering the fund is underfunded, we can define the aim of a given firm as minimizing the difference between fund assets and actuarial liability to an acceptable level. In case the fund is over-funded, minimizing this difference is not accepted. This can be linked to situations that can go without some investment opportunities that could yield potential high profits. It is typical in the literature on the best administration of defined benefit pension systems to take into account quadratic loss functions. The authors proposed a modified game, based on the existing games, for the problem with two different scenarios. Their findings include that Decentralized interaction results in an effective fund excess split. In addition, if the sponsoring firm is worried about the safety of the pension plan, this behavior corresponds in equilibrium with a firm that wants to maximize its own discounted expected utility in the first scenario game. In the first proposed scenario game, if the sponsoring firm is concerned about the pension plan's security, this conduct equilibrates with a firm that seeks to maximize its own discounted anticipated utility.

In [44], the authors proposed addressed the problem of portfolio selection in financial investments. The authors framed this problem as an optimization problem and then they utilized a collaborative neuro-dynamic method that optimized using the particle swarm optimization (PSO) algorithm to select the optimal portfolio. The optimization model includes two contradicting objectives, namely, minimizing the risk value and maximizing the potential portfolio outcome. The authors evaluated their proposed method in four markets, where the proposed method successfully converged in all of them.

The authors in [45] addressed the problem of designing an option portfolio. They proposed a dynamic option portfolio design; they proposed to adjust the design on daily basis. In this context, they proposed using the delta-gamma approximation of the function of the option price to address the problem. Their proposed method is thoroughly evaluated and compared with the state-of-the-art methods, and the results show that their proposed method delivers the performance of the state-of-the-art methods but with much more run-time; the run-time of their method is better than the other methods of comparison by an order of magnitude.

From the discussed literature, we can conclude that there is a lack of datasets collected from the Egyptian insurance market. In addition, ensemble learning models are not explored in the current efforts to build insolvency predictive models. Of note, a comparison between classic machine learning and ensemble learning predictive models does not exist in the current literature. This comparison should outline the performance gap between these two types of models.

III. DATA AND METHODOLOGY

A. DATASET

The introduced dataset represents the financial factors of 11 Egyptian insurance companies over a period of many years. Each company data is presented by 22 numerical values of real data type. Each value represents a financial factor in terms of a ratio. The target value is the solvency rate, a real number as well. Thus, the task of insolvency prediction can be framed as a regression problem with 21 input features and one outcome (i.e., solvency ratio). If the predicted solvency ratio falls out of a predefined range, then the insurance company will suffer from insolvency and vice versa.

The collected data of this work are obtained from the items of the annual financial statements of Egyptian property-liability insurance companies that are published in the annual reports of the Egyptian insurance companies issued by the Egyptian Financial Supervisory Authority. All of these reports are publicly available.[2] This study covers the period between 1999 and 2019. The dataset has one public sector company and ten private sector companies. Table 1 lists the insurance companies' details (i.e., company name, company sector, and the number of data years).

TABLE 1. Details of the insurance companies.

NO.	Company Name	Company sector	No. years data
1	Misr Insurance Holding Company	General	1999-2019
2	Suez Canal Insurance Company	Private	1999-2019
3	Mohandes Insurance Company	Private	1999-2019
4	Delta Insurance Company	Private	1999-2019
5	AIG Insurance Company	Private	1999-2019
6	AMIG Insurance Company	Private	2004-2019
7	ECGE Insurance Company	Private	2004-2019
8	ACE Insurance Company	Private	2004-2019
9	Royal Insurance Company	Private	2004-2019
10	Allianz Insurance Company	Private	2004-2019
11	ESIH Insurance Company	Private	2004-2019

The Insurance Regulatory Information System (IRIS) is a system initially designed by McKinsey and Company. It was developed by the National Association of Insurance Commissioners (NAIC) to provide an early warning system for insurance company insolvency based upon accounting ratio values derived from the annual financial statements of insurance companies [46]. In this study, we used 21 financial ratios that are used to measure efficiency in financial performance and were used primarily by regulators to predict the solvency ratio of an insurance company. Their application to direct insurance companies in the Egyptian insurance market was to predict the insolvency ratio for insurance companies, which was counted from the annual financial statements of Egyptian property-liability insurance companies. Table 2 shows all the ratios that are used with an "acceptable" range of ratios, and Fig. 1 shows the distribution of the solvency ratio of the Egyptian insurance companies in the proposed dataset. In Table 2, each ratio is represented by the letter "R" and the ratio number, where R22 is the output variable. In Fig. 1, the x-axis represents the solvency ratio, whereas the y-axis represents the number of companies that produced the indicated solvency ratio on the x-axis at any year.

In the correlation analysis, Fig. 2 depicts the pairwise correlation (measured using Pearson correlation coefficient) of the utilized independent variable. In other words, Fig. 2 reveals how features/factors are correlated to each other. This information can be used to recognize the most significant financial factors that help in predicting insolvency. The closer the value to zero the less the correlation and vice versa. The target feature, in Fig. 2 represents the outcome variable. From Fig. 2, we can conclude that the target variable is most correlated to feature R17 by a correlation score of 0.8. Then, the second most correlated feature is feature R5 and R21

[2] The source of data.

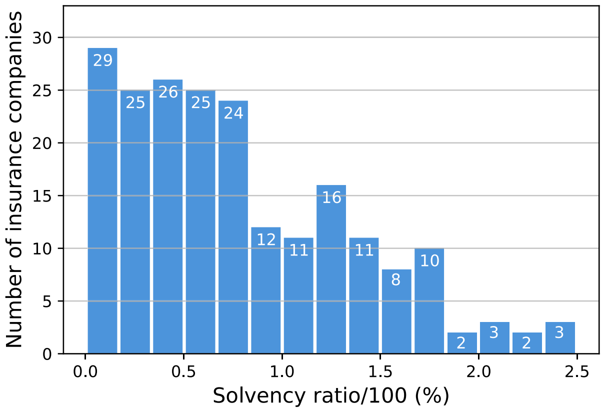

FIGURE 1. **Histogram chart of the values of solvency ratio in the dataset.**

with correlation scores of 0.7 and 0.5, respectively. Thus, the correlation analysis revealed a strong relationship between the outcome feature on one side, and the R17, R5, and R21 ratios on the other side. These three ratios represent the technical provisions to shareholders' funds, insurance companies' debit balances to shareholders, and earnings after taxes to shareholders' funds, respectively.

B. SYSTEM OVERVIEW

The proposed predictive system can be better realized with the help of a block diagram graphical illustration, as in Fig. 3. This is because the proposed system can be understood as a pipeline of well-known machine tasks which were adapted to be suitable for the task of insolvency prediction.

The first phase of the proposed predictive system includes data preparation, as depicted in Fig. 3. In this context, any machine learning model begins with data collection (as described in Sec. III-A). Next, in the pre-processing stage, data must be processed before any future operations can be fed into the machine learning model, the first stage of Fig. 3 as well. Such cleaning tasks often include removing outliers and imputing missing values. Additionally, a feature engineering method (i.e., label encoding) is used to convert categorical variables (i.e., features) values into unique numbers.

Furthermore, in order to assess the performance of the proposed machine learning models, for model validation, we used two validation methods, namely, hold-out validation and k-fold cross-validation. The validation method is a statistical method used to estimate the performance (or accuracy) of machine learning models. The holdout method is a basic and simple approach in which we divide our entire data set into two parts: training data and testing data. In this approach, the data is first shuffled randomly before being split. As the name implies, we train the model on training data and then evaluate it on the testing set. Cross-validation (as depicted in Fig. 4) is used to avoid overfitting a prediction model, which is particularly useful when the amount of data available is limited. The K-fold cross-validation is one strategy for improving the holdout method. In cross-validation, a fixed number of folds (or partitions) of the data is set, then the experiment runs the analysis on each fold. Finally, we calculate the mean and the standard deviation of the overall error estimate.

In the modeling stage of Fig. 3, we used a support vector machine and three ensemble learning machines (i.e., Bagging, Boosting, and Random Forest). Finally, in the fourth stage, we evaluate the performance of the proposed machine learning models with respect to three evaluation metrics (as discussed in Sec. IV-B).

C. METHODOLOGY

Customizing the model's hyperparameters is one of the most challenging aspects of machine learning model design. This is because varying hyperparameter values can result in varying degrees of precision. Our research employed two distinct modes of learning: classical machine learning and ensemble learning. The proposed model makes use of classical machine learning techniques such as SVR and Random Forest.

Each learning model is subjected to a grid search in order to obtain the best parameter tuning. The hyperparameters are critical because they have a significant influence on how a machine learning model behaves in general. The Random Forest regressor allows for the adjustment of parameters such as n_estimators, max_features, max_depth, min_samples_split, min_samples_leaf, and bootstrap. Grid search was used to determine the best value for n_estimators.

The hyperparameters of ensemble learning via Boosting algorithms are critical in resolving the trade-off between bias and variance. For example, when using the Catboost regressor, the number of iterations, learning_rate, leaf_regularization, and tree depth are adjusted to achieve the highest accuracy. By utilizing grid search, we were able to acquire the best value for the number of iterations. Unlike bagging algorithms, which control only the model's high variance, boosting algorithms control both aspects (bias and variance) and are therefore considered to be more effective. For example, when using the Bagging regressor, the base_estimator, n_estimators, max_samples, max_features, bootstrap, and bootstrap_features are adjusted to achieve the highest accuracy. Grid search was used to determine the best value for n_estimators. Table 3 shows the best parameter values for the above-mentioned machine learning models. Other utilized models' parameters which are not listed in Table 3 are set to the default parameters values.

In the proposed work, we utilized the CatBoost [47] boosting algorithm, which is a fast implementation of gradient boosting algorithm and successfully handles categorical features, where the Egyptian-companies variable (represented by R_0 in Fig. 2) is the categorical feature in the proposed work.

IV. EXPERIMENTAL RESULTS

A. EXPERIMENTAL SETUP

The experiments were conducted on a computer with an Intel(R) Core (TM) i7-9750H CPU @ 2.60GHz and 16−GB for RAM. The utilized OS is 64-bit Windows 10. The

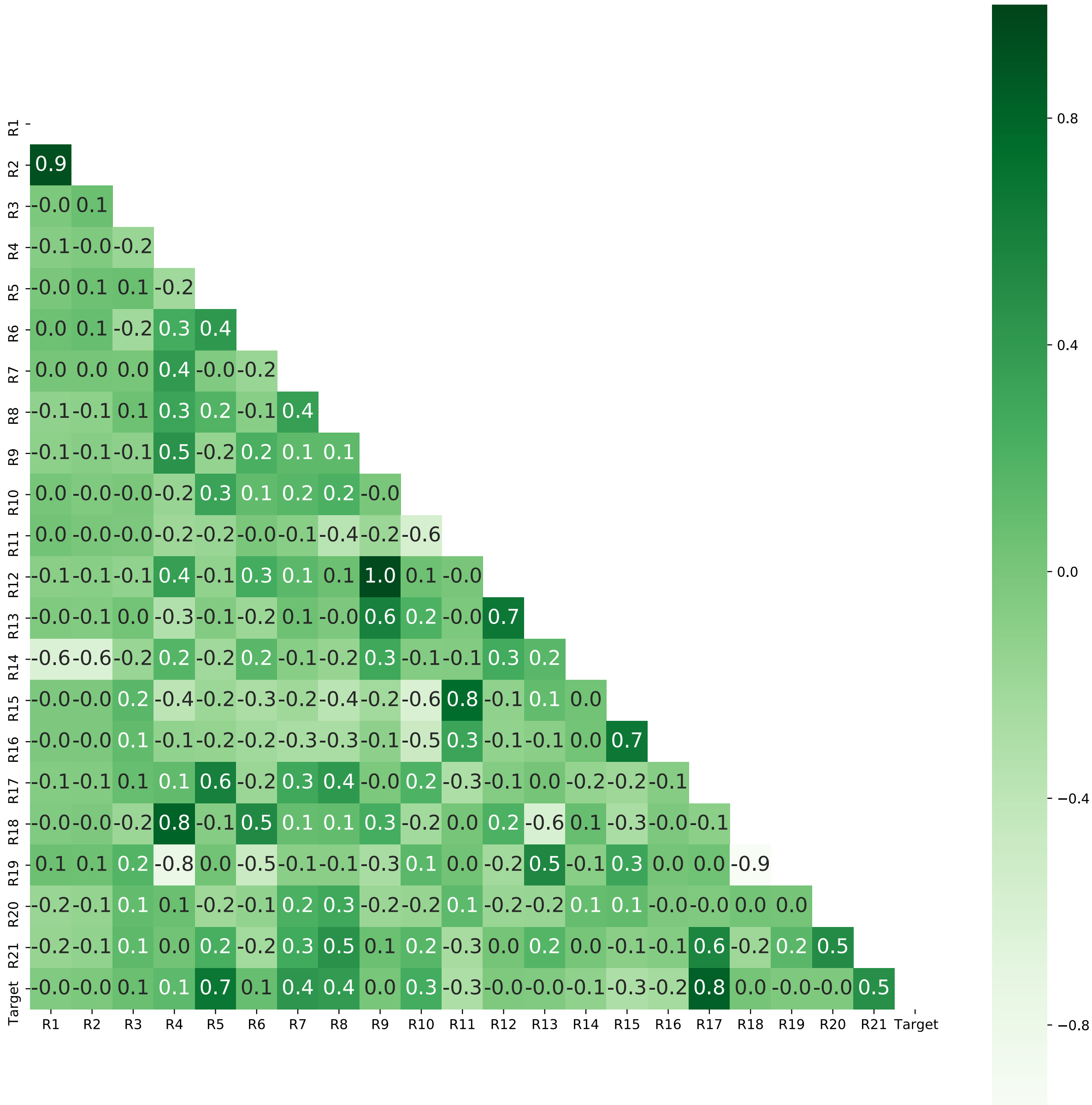

FIGURE 2. **Correlation matrix of the dataset variables.**

Framework implementation is developed in the Python programming language. Moreover, the dataset is loaded using the Pandas [48] data-frame.

Machine learning models are implemented using Scikit-Learn [49] (RF and SVM) and CatBoost[3] libraries. Other commonly used libraries include Matplotlib [50]. To assure the re-producibility of the experimental models, parameter configurations, and reported results, we made the proposed work's source code, visualizations, and data freely accessible online via the author's GitHub page.[4]

For evaluation purposes, we used two validation methods, typically, hold-out validation and k-fold cross-validation (CV) methods [51], [52]. Hold-out validation divides the data into training and test sets, with the test sets being used to evaluate the predictive model's performance. Typically, we used a splitting ratio of 80% and 20%. On the other hand, the CV is used for estimating the generalization performance. Practically, k-fold CV splits data into k equal parts, then, the predictive model is trained for k times. Fig. 4 depicts an example of a 5-fold CV ($k = 5$) method. As shown in Fig. 4, for $i = 1, 2, \ldots, k$, the model is evaluated on the k^{th} fold, while other folds are used for training.

B. EVALUATION METRICS

We employ the well-known used evaluation metrics to compare and evaluate the performance of the proposed machine learning models. The selected evaluation metrics are based on

[3]https://catboost.ai/

[4]https://github.com/Ahmed-Fathalla/Predicting-insolvency-of-insurance-companies-in-Egyptian

TABLE 2. List of ratios.

NO.	Definition	Equation	Acceptable range
R1	Change in net premiums ratio	(Net premiums at the end of the year - Net premiums at the beginning of the year) / Net premiums at the beginning of the year	(-10%:+30%)
R2	Change in total premiums ratio	(Total premiums at the end of the year - Total premiums at the beginning of the year) / Total premiums at the beginning of the year	(-10%:+30%)
R3	Change in total shareholders' funds ratio	(Total shareholders' funds at the end of the year - Total shareholders' funds at the beginning of the year) / Total shareholders' funds at the beginning of the year	(-10%:+50%)
R4	Liquidity ratio	Technical provisions / liquid assets	Less than 100%
R5	Insurance companies debit balances to shareholders' funds ratio	Insurance companies debit balances / total shareholders' funds ratio	Less than 30%
R6	Insurance companies debit balances to total assets ratio	Insurance companies debit balances / Total assets	Less than 10%
R7	Retention ratio	Net earned premiums / Total earned premiums	More than 50%
R8	Investment yield ratio	Net investment income / Total investments	More than 8%
R9	Loss ratio	Claims incurred /earned premiums	Less than 70%
R10	Commissions and production costs ratio	Commissions and production costs / Total earned premiums	Less than 20%
R11	General administrative expenses ratio	General administrative expenses / Total earned premiums	Less than 10%
R12	Collective ratio	Loss ratio + Commissions and production costs ratio + General administrative expenses ratio	Less than 100%
R13	Operation ratio	Collective ratio – (Total liabilities / liquid assets)	Less than 100%
R14	Profit (Loss) of insurance activities to net investment income ratio	Profit (Loss) of insurance activities / Net investment income	More than 25%
R15	Policyholders' funds and shareholders' funds to net premiums ratio	(Total policyholders' funds + Total shareholders' funds) / Net premiums	More than 150%
R16	Technical provisions to net premiums ratio	Technical provisions / Net premiums	More than 100%
R17	Technical provisions to shareholders' funds	technical provisions / Total shareholders' funds	Less than 350%
R18	Total liabilities to Liquid assets ratio	Total liabilities / Liquid assets	Less than 105%
R19	The shareholders' funds to total assets ratio	(Total shareholders' funds / Total assets)	More than 10%
R20	Earnings before taxes to total assets ratio	Earnings before taxes / Total assets	More than 2%
R21	Earnings after taxes to shareholders' funds ratio	Earnings after taxes / Total shareholders' funds	More than 5%
R22	Solvency ratio	Net earned premiums / Total Shareholders' funds	Less than 200%

TABLE 3. Hyperparameters values of the ML models.

NO.	Model	Parameter	Value
1	Random Forest	n_estimators	100
2	Catboost	iterations	2000
3	Bagging	n_estimators	100

the recently published works on the insolvency of insurance companies' problems [53], [54], [55]. These three metrics include the Mean Absolute Error (*MAE*, Eq. 1), Root Mean Squared Error (*RMSE*, Eq. 2), and Coefficient of determination (R^2, Eq. 3). These metrics represent absolute and relative error evaluation methods.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|y_i - \hat{y}_i\right| \tag{1}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2} \tag{2}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \tag{3}$$

TABLE 4. Proposed models' performance on test set.

Model	MAE	$RMSE$	R^2
Bagging	0.0706	0.1060	96.88%
CatBoost	**0.0518**	**0.0914**	**97.68%**
Random forest	0.0737	0.1130	96.45%
SVR	0.2985	0.4119	52.96%
Decision Tree	0.1333	0.2244	86.04%
Linear Regression	0.1908	0.236	84.56 %
Lasso	0.4442	0.532	21.53%
Ridge	0.2351	0.4061	54.28%
KNN	0.0819	0.1297	95.34%

where N denotes the number of observations, y_i indicates true values, $\hat{y}_i$ signifies predicted values, and $\bar{y}$ specifies the mean true values.

C. RESULTS

The evaluation of the proposed models is conducted over the model's accuracy and CV (i.e., variance). In the accuracy

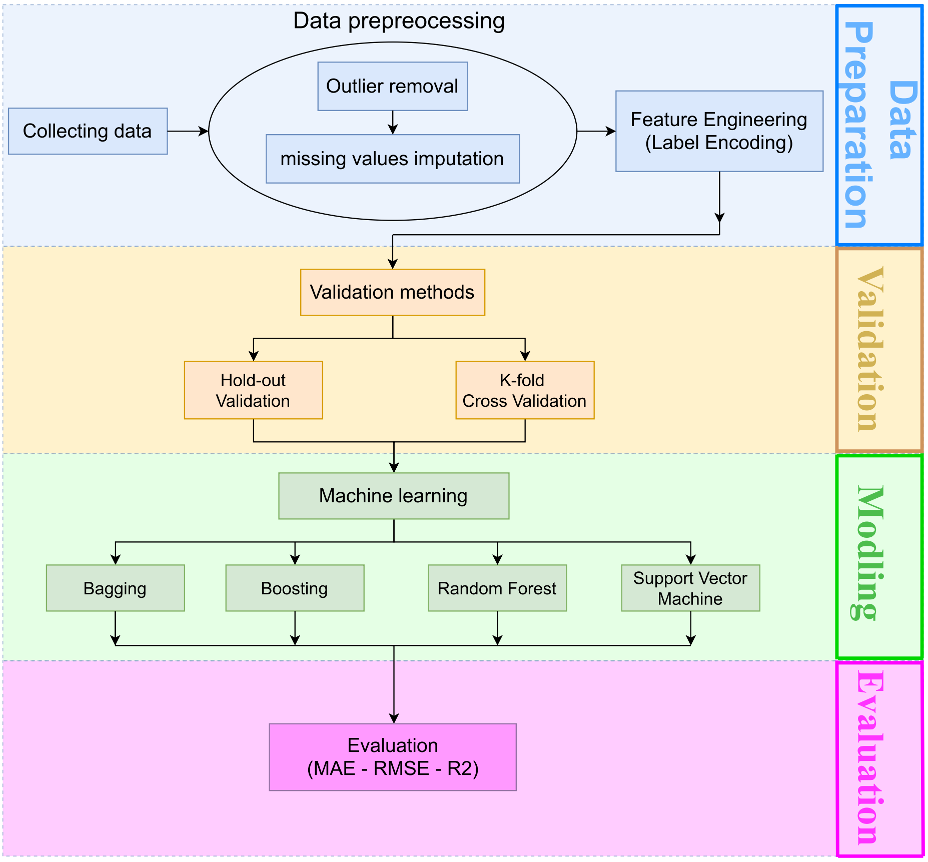

FIGURE 3. Block diagram of the proposed method.

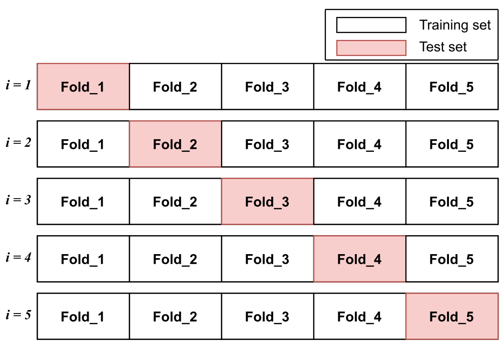

FIGURE 4. Cross-validation method.

evaluation, the proposed models will be examined over two types of metrics, namely, the error measure (i.e., MAE and RMSE) and the measurement can the variation of a dependent variable be explained by the independent variable(s), i.e., R^2. In the second point of evaluation, the variance of the performance of the proposed model will be examined using k-fold cross-validation.

Table 4 lists the scores of the three metrics for the first point of comparison, i.e., the accuracy. The methods of comparison include two ensemble learning methods and seven classic machine learning methods. Obviously, the CatBoost model outperforms the other models on all of the three metrics, while the SVR and Lasso models produced the worst performance. The Lasso performance is expected as its best results are achieved when the number of features is larger than the number of training samples [56], which is not the case with the proposed dataset. The proposed dataset has 21 input features with a few hundred training samples. Of note, the bad performance of the SVR model can be linked to the fact that the data is imbalanced [57]. The collected data covered a different number of years, based on the company, as shown in Table 1.

For the second point of evaluation, the proposed models' performance variance is listed in Table 5 for 15-fold cross-validation. Despite changing the data split ratio between the train and test sets, the performance gaps between the four proposed models are still the same as reported in Table 4. The reported standard deviation of the three accuracy metrics is small compared to the average accuracy values, except for the support vector machine model, which has a relatively large variance. For the R^2 metrics, Fig. 5 depicts the variance over the 15-folds. Of note, the bagging model has the least reported variance and is then followed by the random forest model while the CatBoost model has the most variance. Similarly, Fig. 6 depicts the variance in the MAE metric for the random forest, bagging, and CatBoost models. The performance

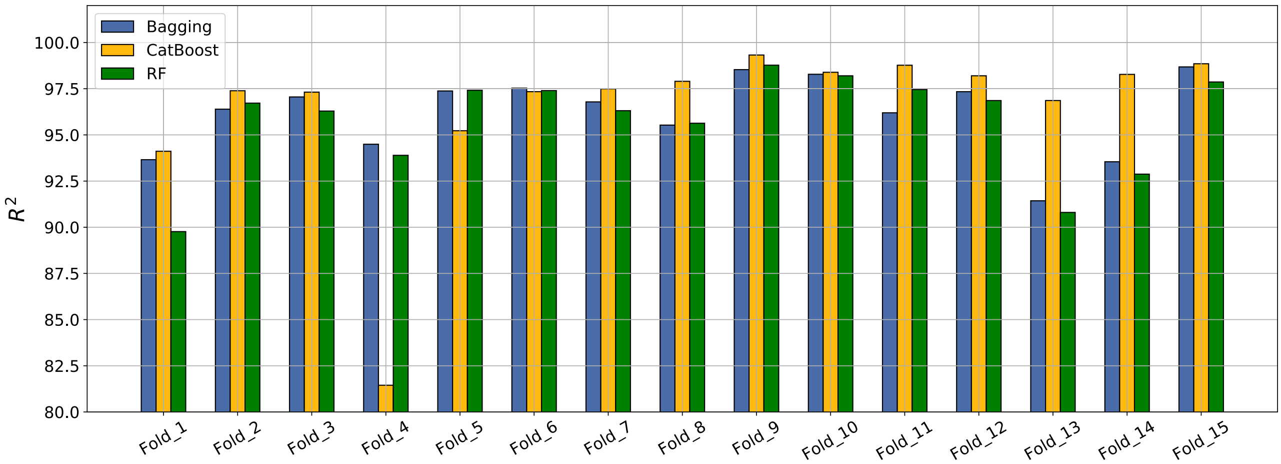

FIGURE 5. R^2 **score of 15-fold cross-validation.**

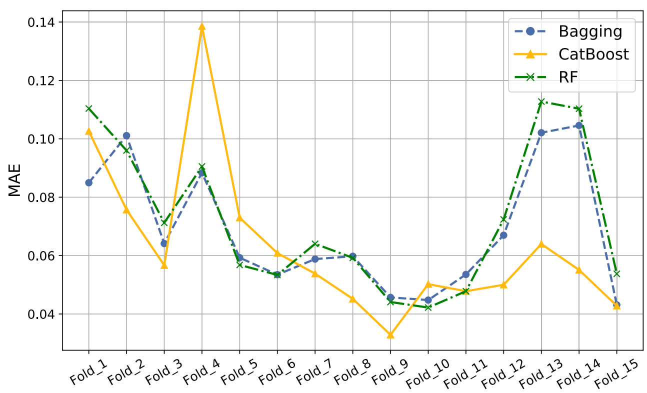

FIGURE 6. **MAE of 15-fold cross-validation.**

TABLE 5. **Proposed models' performance using 15-fold cross-validation.**

Model	MAE	$RMSE$	R^2
Bagging	0.0687 ± 0.0201	0.1089 ± 0.0302	96.19% ± 1.93%
CatBoost	**0.0633 ± 0.0246**	**0.0999 ± 0.0521**	**96.46% ± 4.04%**
Random forest	0.0723 ± 0.0233	0.1139 ± 0.0349	95.75% ± 2.51%
SVR	0.2979 ± 0.0409	0.3976 ± 0.0768	51.61% ± 12.85%
Decision Tree	0.098 ± 0.0383	0.1534 ± 0.0779	91.58% ± 8.94%
Linear Regression	0.2284 ± 0.0688	0.2974 ± 0.1140	67.52% ± 35.58%
Lasso	0.4337 ± 0.0348	0.5286 ± 0.0564	15.08% ± 8.15%
Ridge	0.2199 ± 0.0691	0.2869 ± 0.0958	71.35% ± 23.49%
KNN	0.1265 ± 0.057	0.1928 ± 0.0922	86.15 ± 14.32 %

variance of these three models is identical to the reported variance of Fig. 5.

V. CONCLUSION

In this work, we proposed collecting the first insolvency dataset for Egyptian insurance companies. The collected dataset is spread over 20 years and includes 21 important financial ratios alongside the solvency ratio (i.e., the target variable). Then, this proposed dataset is utilized for training a predictive model of insolvency in insurance companies. These proposed predictive models are developed based on two different machine learning techniques, namely, 1) classic machine learning models such as SVM and RF, and 2) ensemble learning models such as bagging and gradient boosting on decision trees. Next, these two techniques are compared for the sake of studying the performance gap between them. To evaluate the proposed models, a cross-validation analysis is performed and several accuracy metrics are measured. The experimental results outlined that the ensemble learning models outperformed the classic machine learning models. The best predictive model for insolvency from the Egyptian insurance companies' data is the gradient boosting on decision trees.

ACKNOWLEDGMENT

The authors would like to acknowledge with much appreciation Prince Sattam Bin Abdulaziz University for its ongoing support of our research.

REFERENCES

[1] X. Xu and Y. Wang, "Financial failure prediction using efficiency as a predictor," *Expert Syst. Appl.*, vol. 36, no. 1, pp. 366–373, Jan. 2009.
[2] A. French, M. Vital, and D. Minot, "Insurance and financial stability," Bank England Quart. Bull., London, U.K., Tech. Rep. Q3/3, 2015, pp. 242–257, vol. 3.
[3] Z. Rustam and G. Saragih, "Prediction insolvency of insurance companies using random forest," *J. Phys., Conf. Ser.*, vol. 1752, no. 1, Feb. 2021, Art. no. 012036.
[4] A. Ibiwoye, O. Ajibola, and A. B. Sogunro, "Artificial neural network model for predicting insurance insolvency," *Int. J. Manag. Bus. Res.*, vol. 2, no. 1, pp. 59–68, 2012.
[5] A. A. Khalil, Z. Liu, and A. A. Ali, "Using an adaptive network-based fuzzy inference system model to predict the loss ratio of petroleum insurance in Egypt," *Risk Manage. Insurance Rev.*, vol. 25, no. 1, pp. 5–18, Apr. 2022.
[6] J. D. Cummins, M. F. Grace, and R. D. Phillips, "Regulatory solvency prediction in property-liability insurance: Risk-based capital, audit ratios, and cash flow simulation," *J. Risk Insurance*, vol. 66, pp. 417–458, Sep. 1999.
[7] R. Bar-Niv and M. L. Smith, "Underwriting, investment and solvency," *J. Insurance Regulation*, vol. 5, no. 4, p. 409, 1987.

[8] R. BarNiv, "Identifying financial distress in the insurance industry: A methodology inquiry," in *Proc. Annu. Meeting Amer. Risk Insurance Assoc.*, 1989.

[9] P. J. Harmelink, "Prediction of best's general policyholders' ratings," *J. Risk Insurance*, vol. 41, no. 4, pp. 621–632, Dec. 1974.

[10] J. M. Ambrose and A. M. Carroll, "Using best's ratings in life insurer insolvency prediction," *J. Risk Insurance*, vol. 61, no. 2, pp. 317–327, Jun. 1994.

[11] J. D. Stowe and C. J. Watson, "A multivariate analysis of the composition of life insurer balance sheets," *J. Risk Insurance*, vol. 52, no. 2, pp. 222–240, Jun. 1985.

[12] J. S. Trieschmann and G. E. Pinches, "A multivariate model for predicting financially distressed PL insurers," *J. Risk Insurance*, vol. 40, no. 3, pp. 327–338, Sep. 1973.

[13] P. L. Brockett, W. W. Cooper, L. L. Golden, and U. Pitaktong, "A neural network method for obtaining an early warning of insurer insolvency," *J. Risk Insurance*, vol. 61, no. 3, pp. 402–424, Sep. 1994.

[14] P. L. Brockett, W. W. Cooper, L. L. Golden, and X. Xia, "A case study in applying neural networks to predicting insolvency for property and casualty insurers," *J. Oper. Res. Soc.*, vol. 48, no. 12, pp. 1153–1162, Dec. 1997.

[15] M. J. S. Vargas, J. A. G. Fana, A. J. H. Martínez, J. L. V. Zanón, and A. S. Arellano, "Using rough sets to predict insolvency of Spanish non-life insurance companies," Facultad de Ciencias Económicas y Empresariales, Universidad Complutense de Madrid, Madrid, Spain, Tech. Rep. 03-02, 2003.

[16] P. L. Brockett, L. L. Golden, J. Jang, and C. Yang, "A comparison of neural network, statistical methods, and variable choice for life insurers' financial distress prediction," *J. Risk Insurance*, vol. 73, no. 3, pp. 397–419, Sep. 2006.

[17] S.-H. Hsiao and T.-J. Whang, "A study of financial insolvency prediction model for life insurers," *Expert Syst. Appl.*, vol. 36, no. 3, pp. 6100–6107, Apr. 2009.

[18] N. S. Chiet, S. H. Jaaman, N. Ismail, and S. M. Shamsuddin, "Insolvency prediction model using artificial neural network for Malaysian general insurers," in *Proc. World Congr. Nature Biologically Inspired Comput. (NaBIC)*, 2009, pp. 584–589.

[19] C. Lin, "Using neural networks as a support tool in the decision making for insurance industry," *Expert Syst. Appl.*, vol. 36, no. 3, pp. 6914–6917, Apr. 2009.

[20] P. Shuang and L. Wei, "The early warning of life insurance company based on BP artificial neural network," in *Proc. Int. Conf. Inf. Manage. Eng.* Berlin, Germany: Springer, 2011, pp. 30–38.

[21] E. A. Bakr and O. Hanafy, "Using artificial neural network to predict the insolvency for Egyptian insurance companies," *J. Financial Commercial Stud.*, vol. 2, pp. 132–153, Jan. 2013.

[22] Z. Rustam and F. Yaurita, "Insolvency prediction in insurance companies using support vector machines and fuzzy kernel C-means," *J. Phys., Conf. Ser.*, vol. 1028, Jun. 2018, Art. no. 012118.

[23] D. Amodei *et al.*, "Deep speech 2: End-to-end speech recognition in English and Mandarin," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 173–182.

[24] W. Chan, N. Jaitly, Q. V. Le, O. Vinyals, and N. M. Shazeer, "Speech recognition with attention-based recurrent neural networks," U.S. Patent 9 990 918, Jun. 5, 2018.

[25] A. Ali, A. Fathalla, A. Salah, M. Bekhit, and E. Eldesouky, "Marine data prediction: An evaluation of machine learning, deep learning, and statistical predictive models," *Comput. Intell. Neurosci.*, vol. 2021, pp. 1–13, Nov. 2021.

[26] E. Eldesouky, M. Bekhit, A. Fathalla, A. Salah, and A. Ali, "A robust UWSN handover prediction system using ensemble learning," *Sensors*, vol. 21, no. 17, p. 5777, Aug. 2021.

[27] Y. Goldberg, "A primer on neural network models for natural language processing," *J. Artif. Intell. Res.*, vol. 57, pp. 345–420, Nov. 2016.

[28] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[29] Y. Kim, "Convolutional neural networks for sentence classification," 2014, *arXiv:1408.5882*.

[30] A. Vaswani, S. Bengio, E. Brevdo, F. Chollet, A. N. Gomez, S. Gouws, L. Jones, Ł. Kaiser, N. Kalchbrenner, N. Parmar, R. Sepassi, N. Shazeer, and J. Uszkoreit, "Tensor2Tensor for neural machine translation," 2018, *arXiv:1803.07416*.

[31] M. Nikou, G. Mansourfar, and J. Bagherzadeh, "Stock price prediction using DEEP learning algorithm and its comparison with machine learning algorithms," *Intell. Syst. Accounting, Finance Manage.*, vol. 26, no. 4, pp. 164–174, Oct. 2019.

[32] A. Fathalla, A. Salah, K. Li, K. Li, and P. Francesco, "Deep end-to-end learning for price prediction of second-hand items," *Knowl. Inf. Syst.*, vol. 62, no. 12, pp. 4541–4568, Dec. 2020.

[33] Z. Alameer, A. Fathalla, K. Li, H. Ye, and Z. Jianhua, "Multistep-ahead forecasting of coal prices using a hybrid deep learning model," *Resour. Policy*, vol. 65, Mar. 2020, Art. no. 101588.

[34] Z. Fan and Y. Wang, "Stock price prediction based on deep reinforcement learning," in *Innovative Computing*. Singapore: Springer, 2022, pp. 845–852.

[35] E. G. Dada, J. S. Bassi, H. Chiroma, S. M. Abdulhamid, A. O. Adetunmbi, and O. E. Ajibuwa, "Machine learning for email spam filtering: Review, approaches and open research problems," *Heliyon*, vol. 5, no. 6, Jun. 2019, Art. no. e01802.

[36] M. E. Abbas, Z. Chengzhang, A. Fathalla, and Y. Xiao, "End-to-end antigenic variant generation for H1N1 influenza HA protein using sequence to sequence models," *PLoS ONE*, vol. 17, no. 3, Mar. 2022, Art. no. e0266198.

[37] B. T. Grys, D. S. Lo, N. Sahin, O. Z. Kraus, Q. Morris, C. Boone, and B. J. Andrews, "Machine learning and computer vision approaches for phenotypic profiling," *J. Cell Biol.*, vol. 216, no. 1, pp. 65–71, Jan. 2017.

[38] W. Lin, Z. Wu, L. Lin, A. Wen, and J. Li, "An ensemble random forest algorithm for insurance big data analysis," *IEEE Access*, vol. 5, pp. 16568–16575, 2017.

[39] M. A. Fauzan and H. Murfi, "The accuracy of XGBoost for insurance claim prediction," *Int. J. Adv. Soft Comput. Appl.*, vol. 10, no. 2, pp. 159–171, 2018.

[40] R. Jain, J. A. Alzubi, N. Jain, and P. Joshi, "Assessing risk in life insurance using ensemble learning," *J. Intell. Fuzzy Syst.*, vol. 37, no. 2, pp. 2969–2980, Sep. 2019.

[41] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, 1996.

[42] R. E. Schapire, "The strength of weak learnability," *Mach. Learn.*, vol. 5, no. 2, pp. 197–227, Jul. 1990.

[43] R. Josa-Fombellida and J. P. Rincón-Zapatero, "Equilibrium strategies in a defined benefit pension plan game," *Eur. J. Oper. Res.*, vol. 275, no. 1, pp. 374–386, 2019.

[44] M.-F. Leung and J. Wang, "Minimax and biobjective portfolio selection based on collaborative neurodynamic optimization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 7, pp. 2825–2836, Jul. 2021.

[45] L. Zhao and D. P. Palomar, "A Markowitz portfolio approach to options trading," *IEEE Trans. Signal Process.*, vol. 66, no. 16, pp. 4223–4238, Aug. 2018.

[46] *Iris Ratios Manual for Property/Casualty, Life/ Accident & Health, and Fraternal, Insurance Regulatory Information Systems (IRIS) Manual*, Nat. Assoc. Insurance Commissioners, Kansas City, MI, USA, 2017.

[47] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: Unbiased boosting with categorical features," 2017, *arXiv:1706.09516*.

[48] W. McKinney, "Data structures for statistical computing in Python," in *Proc. 9th Python Sci. Conf.*, vol. 445. Austin, TX, USA, 2010, pp. 51–56.

[49] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, A. Müller, J. Nothman, G. Louppe, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Oct. 2012.

[50] J. D. Hunter, "Matplotlib: A 2D graphics environment," *Comput. Sci. Eng.*, vol. 9, no. 3, p. 90, 2007.

[51] M. Stone, "Cross-validatory choice and assessment of statistical predictions," *J. Roy. Stat. Soc. B, Methodol.*, vol. 36, no. 2, pp. 111–133, Jan. 1974.

[52] A. Blum, A. Kalai, and J. Langford, "Beating the hold-out: Bounds for K-fold and progressive cross-validation," in *Proc. 12th Annu. Conf. Comput. Learn. Theory (COLT)*, 1999, pp. 203–208.

[53] J.-Y. Lin, C.-T. Cheng, and K.-W. Chau, "Using support vector machines for long-term discharge prediction," *Hydrolog. Sci. J.*, vol. 51, no. 4, pp. 599–612, Aug. 2006.

[54] D. Ashourloo, H. Aghighi, A. A. Matkan, M. R. Mobasheri, and A. M. Rad, "An investigation into machine learning regression techniques for the leaf rust disease detection using hyperspectral measurement," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 9, pp. 4344–4351, Sep. 2016.

[55] L. Wang, O. Kisi, B. Hu, M. Bilal, M. Zounemat-Kermani, and H. Li, ‘‘Evaporation modelling using different machine learning techniques,’’ *Int. J. Climatol.*, vol. 37, pp. 1076–1092, Aug. 2017.
[56] M. Yamada, W. Jitkrittum, L. Sigal, E. P. Xing, and M. Sugiyama, ‘‘High-dimensional feature selection by feature-wise kernelized Lasso,’’ *Neural Comput.*, vol. 26, no. 1, pp. 185–207, 2014.
[57] W.-Z. Lu and D. Wang, ‘‘Ground-level ozone prediction by support vector machine approach with a cost-sensitive classification scheme,’’ *Sci. Total Environ.*, vol. 395, nos. 2–3, pp. 109–116, 2008.

AHMED A. KHALIL received the M.Sc. degree in insurance from Assiut University, Assiut, Egypt, in 2019. He is currently pursuing the Ph.D. degree with the School of Mathematics and Statistics, Central South University, Hunan, China. His research interests include insurance, risk management, statistical learning, and machine learning.

ZAIMING LIU was the Former Dean of the School of Mathematics and Statistics and the Former Secretary of the Party Committee. He is currently a Professor with the School of Mathematics and Statistics, and the Doctoral Supervisor of the Central South University, Hunan, China. His research interests include queuing theory and queuing networks, insurance risk theory, and stochastic process and its application.

AHMAD SALAH received the master’s degree in CS from Ain-shams University, Cairo, Egypt, and the Ph.D. degree in computer science from Hunan University, China, in 2014. He is currently an Associate Professor of computer science and technology with Zagazig University, Egypt. He has published more than 40 papers in peer-reviewed journals, such as the IEEE Transactions on Parallel and Distributed Systems, IEEE-ACM Transactions on Computational Biology and Bioinformatics, and *ACM Transactions on Parallel Computing*. His current research interests include parallel computing, computational biology, and machine learning.

AHMED FATHALLA received the Ph.D. degree in computer science and technology Hunan University, Changsha, China, in 2021. His research interest includes machine learning.

AHMED ALI received the degree in computer science from the College of Computers and Informatics, Suez Canal University, Ismailia, Egypt, in 2005, and the M.Sc. and Ph.D. degrees from the College of Information Science and Engineering, Hunan University, in 2011 and 2016, respectively. He is currently working as an Assistant Professor with the Department of Computer Science, College of Computer Engineering and Sciences, Prince Sattam Bin Abdulaziz University, Al-Kharj, Saudi Arabia. His research interests include machine learning, signal processing, channel estimation and modulation, PAPR, the IoT, and VANETs.