Employee Attrition Prediction using Nested Ensemble Learning Techniques

Muneera Saad Alshiddy, Bader Nasser Aljaber
Information Systems Department, College of Computer and Information Sciences
Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, Kingdom of Saudi Arabia

Abstract—**In many industries, including the IT industry, rising employee attrition is a major concern. Hiring a candidate for an unsuitable job because of issues with the employment process can lead to employee attrition. Thus, enhancing the employment process would reduce the attrition rate. This paper aims to investigate the effect of ensemble learning techniques on enhancing the employment process by predicting employee attrition. This paper applied a two-layer nested ensemble model to the *IBM HR Analytics Employee Attrition & Performance* dataset. The performance of this model was compared to that of the random forest (RF) algorithm as a baseline for comparison. The results showed that the proposed model outperformed the baseline algorithm. The RF model achieved an accuracy of 94.2417%, an F1-score of 94.2%, and an AUC of 98.4%. However, the proposed model had the highest performance. It outperformed with an accuracy of 94.5255%, an F1-score of 94.5%, and an AUC of 98.5%. The performance of the proposed model was compared with that of the baseline comparison algorithm by using a paired t-test. According to the paired t-test, the performance of the proposed model was statistically better than that of the baseline comparison algorithm at the significance level of 0.05. Thus, the two-layer nested ensemble model improved the employee attrition prediction.**

Keywords*—*Nested ensemble learning; employee attrition; machine learning; employment process

I. Introduction

The workforce is a crucial asset of an organization due to the strong positive correlation between the success of the organization and its employees' performance. The performance of employees can be affected by human resource management practices, such as the selection process [1], which aims to choose suitable candidates for specific job vacancies based on job-related criteria [2]. The wrong selection of candidates can lead to something called employee attrition, which is described as a large decrease in the labor force. Employee attrition occurs for several reasons [3], such as professional reasons, personal reasons, workplace challenges, and poor employee-job fit [4]. The last reason has a significant relationship with the employment process. It is about hiring a candidate for an inappropriate job because of issues in the employment process [5], or, in other words, the wrong selection of candidates. While a poor employment process can increase the chance of employee attrition, taking steps to reduce the rate of attrition can improve the employment process. Thus, employee attrition can be considered an indicator of the effectiveness of the employment process.

Some literature has studied employee attrition using machine learning (ML) techniques including but not limited to the works [6]–[15]. They conducted their experiments on the same dataset, the *IBM HR Analytics Employee Attrition & Performance* dataset. Three of them [6], [10], [13] used ensemble techniques in their experiments. To the best of the authors' knowledge, this study is the first to apply a nested ensemble technique to the *IBM HR Analytics Employee Attrition & Performance* dataset.

The problem addressed in this paper is the increasing rate of employee attrition [16], which is a major concern in several sectors, such as the IT industry. Employee attrition impacts an organization's performance. It drains many budgets and affects employee satisfaction [3]. Thus, it is necessary to control the rate of attrition [16]. This paper seeks to find the answer of the following question: What is the effect of using nested ensemble learning techniques to improve employee attrition prediction when applied to an employee dataset that has been optimized by a feature selection technique? Additionally, this study aims to investigate the effect of using nested ensemble learning techniques to improve employee attrition prediction when applied to an employee dataset that has been optimized by a feature selection technique.

The rest of this paper is organized as follows: Section II gives a background on the subject and an overview of the related work. Section III explains the methodology. Section IV presents the results and discussion. Section V summarizes the findings of the research and recommends future research directions.

II. Related Work

This paper used the *IBM HR Analytics Employee Attrition & Performance* dataset. It was developed by IBM's data scientists and made publicly available on the Kaggle website [17]. It was chosen because its attributes can be obtained easily from HR departments and its characteristics can simulate real-world HR issues [18]. Several studies conducted their research on this dataset using ML techniques. The current research adopts some of them as related work and reviews them in this section. Table I shows some information about the related work.

Each study was summarized in a single paragraph by reviewing the methods adopted for balancing data, selecting features, applying algorithms, and evaluating performance. Then, each paragraph was concluded by evaluating the performance of the research model using performance evaluation metrics.

The work [6], data balancing was not applied and was considered for future work. As well, feature selection techniques

TABLE I. A BRIEF OVERVIEW OF THE RELATED WORK

Ref	Paper's Name	Year	Publication's Name	Pub. Type	Pub. Rank
[6]	A Comparison of Machine Learning Approaches for Predicting Employee Attrition	2022	Applied Sciences	Journal	Q3
[7]	Counterfactual Explanation Trees: Transparent and Consistent Actionable Recourse with Decision Trees	2022	International Conference on Artificial Intelligence and Statistics (AISTATS)	Conference	A
[8]	Design of System-of-System Acquisition Analysis Using Machine Learning	2022	Complexity	Journal	Q2
[9]	Predicting Employee Attrition Using Machine Learning Approaches	2022	Applied Science	Journal	Q3
[10]	Talent management by predicting employee attrition using enhanced weighted forest optimization algorithm with improved random forest classifier	2022	International Journal of Advanced Technology and Engineering Exploration	Journal	Q4
[11]	An Improved Machine Learning-Based Employees Attrition Prediction Framework with Emphasis on Feature Selection	2021	Mathematics	Journal	Q2
[12]	Employee attrition estimation using random forest algorithm	2021	Baltic Journal of Modern Computing	Journal	Q4
[13]	From Big Data to Deep Data to Support People Analytics for Employee Attrition Prediction	2021	IEEE Access	Journal	Q1
[14]	Predicting Employee Attrition Using Machine Learning Techniques	2020	Computers	Journal	Q2
[15]	A System for Analysis and Remediation of Attrition	2019	IEEE International Conference on Big Data	Conference	B

were not used. The research built models using several algorithms, such as logistic regression (LR), decision tree (DT), random forest (RF), naive bayes (NB), and artificial neural networks (ANN). It used these algorithms to build a voting ensemble based on the majority vote strategy (hard vote). The performance of the models was measured by accuracy, F1-score, and AUC metrics. The best performance in terms of accuracy and AUC went to LR, which scored 87.96% and 85.01%, respectively. The highest F1-score reached 33.78%, which was accomplished by NB.

The work [7] did not apply data balancing, feature selection methods, or ensemble learning techniques. It applied lightGBM and TabNet algorithms. It evaluated the performance of the models using an accuracy metric. LightGBM had the best performance, with an accuracy of 85.3%.

In [8], neither data balancing nor feature selection methods were mentioned. It did not use ensemble learning techniques. Six ML algorithms were applied, which were LR, DT, SVM, NB, k-nearest neighbors (KNN), and RF. The accuracy metric was used to evaluate the performance of the models. SVM achieved the highest accuracy, with a value of 86.77%.

The work [9] used SMOTE to balance the dataset. It did not use feature selection methods to select appropriate features. Likewise, it did not apply ensemble learning techniques. However, it applied four algorithms named the extra tree classifier (ETC), SVM, LR, and DT. The performance of the models was estimated using four measures: accuracy, precision, recall, and F1-score. The performance of ETC was the highest, scoring 93% in all four measures.

The work [10] did not mention handling the imbalanced dataset. However, it applied a filter feature selection technique called information gain. As well, it applied bagging ensemble learning. Eight algorithms were used: RF, NB, LR, SVM, KNN, AdaBoost, DT, and logistic model tree (LMT). The performance was evaluated using four metrics: accuracy, precision, recall, and F1-measure. The performance of LMT was the best in terms of accuracy and recall. It achieved 86.94% and 86.90%, respectively. The performance of RF was the best in terms of precision, which was 85.50%. The performance of SVM was the best in terms of the F1-measure, which was 85.10%.

In [11], data balancing techniques were not mentioned. It used a wrapper feature selection technique named max-out, which was a proposed method. The ensemble learning technique was not applied. This work applied just one algorithm, LR. Accuracy, precision, recall, and F1-score measures were used to evaluate the performance of the models. It had an accuracy of 81%, a precision of 43%, a recall of 82%, and an F1-score of 56%.

In [12], both data balancing techniques and feature selection techniques were not mentioned. The ensemble learning technique was not implemented. Six algorithms were applied to the dataset: classification and regression tree, RF, LR, SVM, KNN, and NB. The performance of the models was assessed using accuracy and AUC. The best performance in terms of accuracy went to RF with 85.12%, while the best performance in terms of AUC went to LR with 80.85%.

Data balancing in [13] was not applied, and it was considered for future work. Two feature selection techniques were used: recursive feature elimination, a wrapper method, and selectKBest, a filter method. It used two ensemble learning techniques: the stacking ensemble learning technique and the voting ensemble technique based on the majority vote strategy (hard vote). Furthermore, eight algorithms were implemented, which were DT, LR, SVM, RF, XGBoost, DNN, long short-term memory (LSTM), and convolutional neural networks (CNN). Accuracy and F1-score were used to evaluate the performance of the models. The voting ensemble technique achieved the highest accuracy with 96%, while RF achieved the highest F1-score with 82.8%.

In [14], data balancing, feature selection methods, and ensemble learning techniques were not applied. It applied eight algorithms, which were Gaussian NB, Bernoulli NB, LR, KNN, DT, RF, SVM, and linearSVM. The performance of the models was evaluated using five metrics: accuracy, precision, specificity, recall, and F1-score. The performance of SVM was the best in terms of accuracy, precision, and specificity. It achieved 87.9%, 87.9%, and 99.4%, respectively. In addition, the performance of Gaussian NB was the best in terms of recall and F1-score. It obtained 54.1% and 44.6%, respectively.

The final work, [15], did not mention handling imbalanced data. Moreover, it did not use feature selection techniques to choose the relevant features. Besides, it did not apply ensemble learning techniques. However, it built CLARA, a system designed to enhance employee retention. It compared

CLARA's performance to six algorithms: RF, XGBoost, SVM, spectral clustering, standalone k-means clustering, and standalone frequent pattern mining. It used precision to evaluate the performance. The performance of the proposed system achieved a precision of approximately 70%.

III. Methodology

The idea of the proposed solution is to apply a nested ensemble model to predict employee attrition. To this end, the research prepared the dataset and used the nested ensemble model. All experiments in this paper were done using Waikato Environment for Knowledge Analysis (WEKA) platform version 3.8.6.

A. The Preprocessing Phase

This section shows the steps that were done to prepare the dataset. The preprocessing phase aims to prepare the dataset to be ready for applying ML algorithms in order to obtain the best results. In this research, different steps were applied to deal with several issues that may affect the ML performance, such as having values that are missing, duplicated, or irrelevant, or having features with various scales.

1) Dataset: The dataset used is the *IBM HR Analytics Employee Attrition & Performance* dataset, a public dataset generated by IBM data scientists and available on Kaggle [17]. It consists of 1470 instances and 35 features. Table II shows a brief description of the dataset. The table outlines the data type of each feature, data measurement scales, and descriptions of the feature's values (the number between brackets indicates the number of employees who share the same value).

The feature *Attrition* is a binary categorical variable that has two values: *No* and *Yes*. The value *No* indicates employees who did not leave the company. The value *Yes* indicates employees who left the company. The dataset contains 1233 on-the-job employees, representing 83.88% of the total employees. As for the rest of the employees, representing 16.12%, they left the company. Thus, the dataset is imbalanced because the minority class represents 16.12% of the entire dataset. For data balancing, the SMOTE technique was used. It was adopted because it has been widely used in similar studies; as well, it was applied in the related work [9].

In this dataset, *Attrition* was the second attribute. It was set as a class attribute by choosing *Attribute as class* command after clicking on the *Edit* button under the *Preprocess* tab. Therefore, it was moved to the end of the dataset as the last attribute.

2) Checking for missing and duplicate values: After checking the information shown in the *Selected attribute* panel, the dataset had no missing values. Moreover, it had no duplicate values after checking the values using the *RemoveDuplicates* filter.

3) Checking for unnecessary attributes: Referring to [19], constant attributes or unique attributes are considered unnecessary attributes that should be dropped. Three features were constant, and one had unique values. The features *EmployeeCount*, *Over18*, and *StandardHours* are constant for all instances. Their respective values were 1, Y, and 80. The feature *EmployeeNumber* has employees' identification codes (IDs), which are unique for all instances. Therefore, these four features were dropped because they did not further the research purpose; in other words, they were unnecessary data [19].

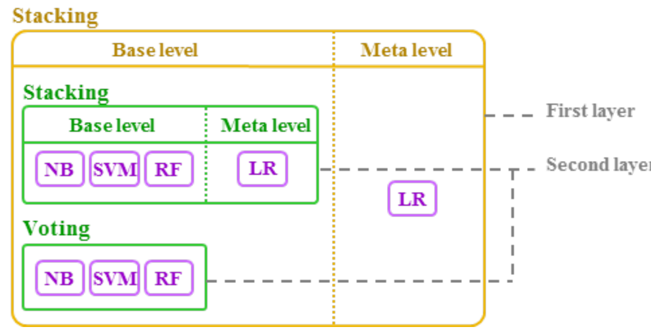

Fig. 1. The proposed model's architecture.

4) Data transformation: The purpose of this step is convert data from one format to another without changing the dataset's content. It helps to improve the ML algorithms' performance and enhance the data's understanding [20]. This research used the *OrdinalToNumeric* filter to unify the data type to be numeric, except *Attrition*, which is the class attribute.

5) Data scaling: Having features with several scales, for instance, the attribute *MonthlyRate* that ranges in thousands and *Age* that ranges in tens, can lead to the *numerical overflow* problem, which is computing numbers that are very small or very big. Thus, it is recommended to put all features on the same scale in order to give them equal weights [21]. This research used the *Normalize* filter to unify all the feature scales to start with 0 and end with 1.

6) The dataset optimization: For optimizing the dataset, the feature selection technique was used, and the information gain-based feature selection method was chosen as it was in the related work [10].

B. The Proposed Model

The proposed model is a nested ensemble learning technique. It is a special type of ensemble learning technique that combines several ensemble methods within one model in order to improve performance [22]. It contains multiple ensemble learning models; each one is inside of another one at different layers. The idea of this model is to use ensemble techniques in the classifier part instead of simple algorithms. In this research, a two-layer nested ensemble model was applied. It used a stacking model in the first layer and both stacking and voting models in the second layer. Moreover, it used traditional algorithms such as NB, SVM, RF, and LR as base-level algorithms of the model's ensemble models. Fig. 1 shows the architecture of the proposed model.

The following ML algorithms were used to build the proposed model:

1) Naive Bayes (NB): It is a probabilistic algorithm that uses probability rules to make predictions [23]. It can deal with different types of data; therefore, it has different models, such as Gaussian, Multinomial, and Bernoulli [24].

TABLE II. THE DATASET DESCRIPTION

No	Attribute	Data Type	Scale	Values Description
1	Age	Numeric	Ratio	Min = 18, Max = 60, Mean = 36.924, StdDev = 9.135
2	Attrition	Textual	Nominal	Yes (237), No (1233)
3	BusinessTravel	Textual	Nominal	Non-Travel (150), Travel_Rarely (1043), Travel_Frequently (277)
4	DailyRate	Numeric	Ratio	Min = 102, Max = 1499, Mean = 802.486, StdDev = 403.509
5	Department	Textual	Nominal	Sales (446), Research & Development (961), Human Resources (63)
6	DistanceFromHome	Numeric	Ratio	Min = 1, Max = 29, Mean = 9.193, StdDev = 8.107
7	Education	Numeric	Ordinal	1 'Below College' (170), 2 'College' (282), 3 'Bachelor' (572), 4 'Master' (398), 5 'Doctor' (48)
8	EducationField	Textual	Nominal	Life Sciences (606), Medical (464), Marketing (159), Technical Degree (132), Human Resources (27), Other (82)
9	EmployeeCount	Numeric	Ratio	1 (1470)
10	EmployeeNumber	Numeric	Nominal	1470 unique identification codes (IDs), start with 1 and end with 2068. They can be classified as the following: IDs $<=$ 500 (377), IDs range between 501 and 1000 (340), IDs range between 1001 and 1500 (348), IDs range between 1501 and 2000 (357), IDs $>$ 2000 (48)
11	EnvironmentSatisfaction	Numeric	Ordinal	1 'Low' (284), 2 'Medium' (287), 3 'High' (453), 4 'Very High' (446)
12	Gender	Textual	Nominal	Female (588), Male (882)
13	HourlyRate	Numeric	Ratio	Min = 30, Max = 100, Mean = 65.891, StdDev = 20.329
14	JobInvolvement	Numeric	Ordinal	1 'Low' (83), 2 'Medium' (375), 3 'High' (868), 4 'Very High' (144)
15	JobLevel	Numeric	Ordinal	1 (543), 2 (534), 3 (218), 4 (106), 5 (69)
16	JobRole	Textual	Nominal	Sales Executive (326), Research Scientist (292), Laboratory Technician (259), Manufacturing Director (145), Healthcare Representative (131), Manager (102), Sales Representative (83), Research Director (80), Human Resources (52)
17	JobSatisfaction	Numeric	Ordinal	1 'Low' (289), 2 'Medium' (280), 3 'High' (442), 4 'Very High' (459)
18	MaritalStatus	Textual	Nominal	Single (470), Married (673), Divorced (327)
19	MonthlyIncome	Numeric	Ratio	Min = 1009, Max = 19999, Mean = 6502.931, StdDev = 4707.957
20	MonthlyRate	Numeric	Ratio	Min = 2094, Max = 26999, Mean = 14313.103, StdDev = 7117.786
21	NumCompaniesWorked	Numeric	Ratio	Min = 0, Max = 9, Mean = 2.693, StdDev = 2.498
22	Over18	Textual	Nominal	Y (1470)
23	OverTime	Textual	Nominal	Yes (416), No (1054)
24	PercentSalaryHike	Numeric	Ratio	Min = 11, Max = 25, Mean = 15.21, StdDev = 3.66
25	PerformanceRating	Numeric	Ordinal	1 'Low' (0), 2 'Good' (0), 3 'Excellent' (1244), 4 'Outstanding' (226)
26	RelationshipSatisfaction	Numeric	Ordinal	1 'Low' (276), 2 'Medium' (303), 3 'High' (459), 4 'Very High' (432)
27	StandardHours	Numeric	Ratio	80 (1470)
28	StockOptionLevel	Numeric	Ordinal	0 (631), 1 (596), 2 (158), 3 (85)
29	TotalWorkingYears	Numeric	Ratio	Min = 0, Max = 40, Mean = 11.28, StdDev = 7.781
30	TrainingTimesLastYear	Numeric	Ratio	Min = 0, Max = 6, Mean = 2.799, StdDev = 1.289
31	WorkLifeBalance	Numeric	Ordinal	1 'Bad' (80), 2 'Good'(344), 3 'Better'(893), 4 'Best' (153)
32	YearsAtCompany	Numeric	Ratio	Min = 0, Max = 40, Mean = 7.008, StdDev = 6.127
33	YearsInCurrentRole	Numeric	Ratio	Min = 0, Max = 18, Mean = 4.229, StdDev = 3.623
34	YearsSinceLastPromotion	Numeric	Ratio	Min = 0, Max = 15, Mean = 2.188, StdDev = 3.222
35	YearsWithCurrManager	Numeric	Ratio	Min = 0, Max = 17, Mean = 4.123, StdDev = 3.568

2) Support Vector Machines (SVM): It is a very powerful and popular algorithm in machine learning. In addition, it works well with complex datasets of small or medium size [25].

3) Random Forests (RF): It is a widely used algorithm consisting of numerous decision trees that learn from subsets of a training dataset [26]. It is a special type of bagging ensemble technique that uses bagging to combine tree models' predictions [27].

4) Stacking ensemble learning technique: Stacking, short for stacked generalization, is an approach that uses different types of algorithms to make their predictions, which they called base-level algorithms. Then, it uses an algorithm to predict the outcome. This algorithm is called a meta-level algorithm [25]. Recently, several studies that applied the stacking ensemble technique have used logistic regression (LR) as a meta-level algorithm [28]. They prefer using LR because of its speed of training and the small number of parameters that it requires [29]. In this research, the algorithms NB, SVM, and RF were used as base-level algorithms, while the algorithm LR was used as a meta-level algorithm.

5) Voting ensemble learning technique: Voting is an approach that uses different types of algorithms to make its prediction. It is similar to the stacking technique except it predicts the final outcome by using the majority voting approach or the averaging approach [25]. In this research, the algorithms NB, SVM, and RF were used as its algorithms. Furthermore, it predicts the final outcome by using the majority voting approach.

C. Performance Evaluation

1) Accuracy: It is a metric that calculates the percentage of correct predictions out of the total number of predictions. It is appropriate to know to what extent the model can predict correctly. Mathematically, accuracy = (TP + TN) / (TP + TN + FP + FN) [30], as shown in Fig. 2.

2) Precision: It is a metric that calculates the percentage of correct predictions of positive instances out of the total number of predictions of positive instances. It is appropriate to know to what extent the model can exclude any instance that actually does not belong to the positive class. Mathematically, precision = TP / (TP + FP) [30], as shown in Fig. 2.

3) Recall: It is a metric that calculates the percentage of correct predictions of positive instances out of the total of actual positive instances. It is appropriate to know to what extent the model can include any instance that actually belongs to the positive class. Mathematically, recall = TP / (TP + FN) [30], as shown in Fig. 2.

4) F1-score: F1-score, or F1-measure, is a metric that measures the harmonic mean of recall and precision [30]. The harmonic mean gives much weight to small values. Hence, a model can get a high F1-score when both recall and precision are high, or it can get a low score when both recall and precision are low [25]. Mathematically, F1-score = 2 * (recall * precision) / (recall + precision) [30], or F1-score = (2 * TP) / (2 * TP + FN + FP) [25], or simply F1-score = (TP + TP) / (TP + FN + TP + FP), as shown in Fig. 2.

5) The Area Under the Curve (AUC): It is a metric that compares the performance of several models for the same dataset. A model with an AUC above 80% is recommended [31]. The higher the AUC, the better the performance [30].

D. Cross-Validation Technique

The cross-validation method is a statistical method that is used to evaluate and compare ML algorithms. Its mechanism is based on dividing datasets into several sets: one is for validating an ML model, and the rest is for learning the ML model. It is called cross-validation because the training set and validation set cross over in sequential rounds to validate every data point [32].

The two most popular cross-validation approaches are *K-fold cross-validation* and *Hold-out validation*. Both of them are effective as long as the dataset is balanced. However, when the dataset is imbalanced, some modifications are applied in order to stratify the folds by the class label. So, the *stratified* keyword would precede the methods' names. The **stratification** process splits the data into folds, and each fold has instances with class labels that are similar to the entire dataset. In this case, every fold represents the class in approximately the correct proportions. Hence, the stratification technique is appropriate for imbalanced datasets [32].

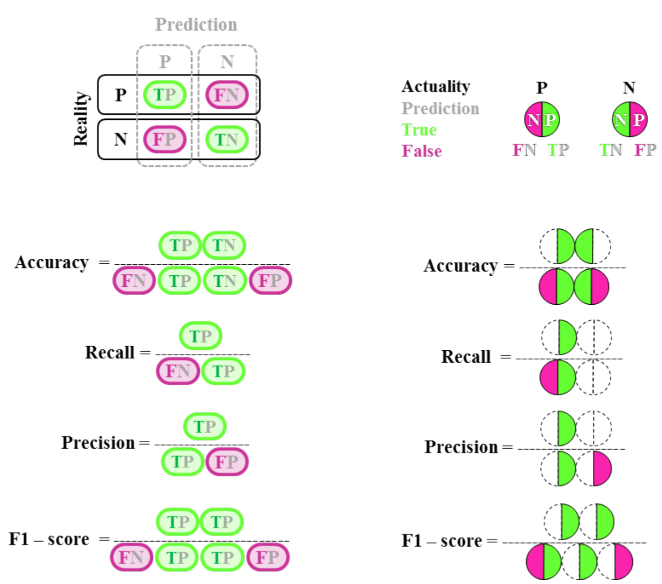

Fig. 2. Illustration of four evaluation metrics derived from the confusion matrix.

Note: The left-hand side clarifies the equations of these metrics, while the right-hand side represents them pictorially.

In this thesis, stratified 10-fold cross-validation was applied because it is recommended for imbalanced datasets [32], [33]. Additionally, the related works [7] and [10] were applied the same technique.

IV. Results and Discussion

This research is part of master's thesis, thus it is limited in time and computer resources. It used the WEKA platform to conduct the research experiments. The scope of this study is limited to the following:

- The research used *IBM HR Analytics Employee Attrition & Performance* dataset because it is a benchmark dataset created by data scientists from IBM, a well-known and reputable company [17]. Further, its features are usually found in real-world employee databases [18]. Besides, it has been widely studied in several research works.
- The research selected its related works on the basis that they used the *IBM HR Analytics Employee Attrition& Performance* dataset, provided they were published in ranked journals or conferences.
- The research studied employee attrition as a criterion to evaluate the effectiveness of the employment process. However, the study did not expand to cover any further criteria.
- The research limited its baseline comparison algorithms to NB, SVM, and RF since they are among the most popular ML techniques [34], especially for binary classification problems [35].
- The research studied nested ensemble techniques by using ensemble models as base-level algorithms. However, the study did not expand to cover any further details of this technique.
- The thesis optimized the dataset by using a feature selection technique.

The proposed model was built as shown in Fig. 1. Its performance was compared with that of NB, SVM, and RF as baseline comparisons, as well as that of other models in related work. The algorithms' hyperparameters were tuned as displayed in Table III. The findings are shown in Table IV.

TABLE III. The Tuning of the Algorithms' Hyperparameters

Algorithm	Hyperparameters' Tuning
NB	useKernelEstimator = True
SVM	C = 2, Kernel = PUK
RF	numIterations = 680

TABLE IV. The Performance of the Baseline Comparison Algorithms and the Proposed Ensemble Model

Algorithm	Accuracy	Precision	Recall	F1-Measure	AUC
NB	90.957%	91.9%	91%	90.9%	95.5%
SVM	92.7818%	92.8%	92.8%	92.8%	92.8%
RF	94.2417%	94.3%	94.2%	94.2%	98.4%
The proposed model	94.5255%	94.5%	94.5%	94.5%	98.5%*

* In terms of the performance evaluated by AUC, the proposed model is statistically better than RF (the baseline) at the significance level of 0.05.

TABLE V. THE PERFORMANCE OF THE MODELS IN THE CURRENT WORK AND THE RELATED WORK

The Work	Algorithm	Accuracy	Precision	Recall	F1-Measure	AUC
[6]	Voting Classifier	79.25%	N/A	N/A	12.22%	83.83%
[7]	LightGBM	85.3%	N/A %	N/A %	N/A %	N/A %
[8]	SVM	86.77 %	N/A %	N/A %	N/A %	N/A %
[9]	Extra Trees	93%	93%	93%	93%	N/A%
[10]	Bagging	83.74%	83.70%	83.70%	77.50%	N/A
[11]	LR	81%	43%	82%	56%	N/A
[12]	RF	85.12%	N/A%	N/A%	N/A%	80.84%
[13]	Voting Classifier	93%	N/A	N/A	58%	N/A
[13]	Stacking	88%	N/A	N/A	50%	N/A
[14]	NB	82.5%	38.6%	54.1%	44.6%	N/A%
[15]	CLARA[1]	65%	N/A%	N/A%	N/A%	N/A%
This Work	The proposed model	94.5255%	94.5%	94.5%	94.5%	98.5%

[1] A proposed end-to-end system.

The results demonstrated that the performance of RF as a baseline comparison algorithm was the best, with an accuracy of 94.2417%, a precision of 94.3%, a recall of 94.2%, an F1-score of 94.2%, and an AUC of 98.4%. However, the proposed model outperformed the other models, including RF. It exceeded the other models in terms of all metrics. It achieved 94.5255% in accuracy, 98.5% in AUC, and 94.5% in precision, recall, and F1-score.

As shown in Table IV, the best performance for each model, except SVM, was given by the AUC. The performance scores given by accuracy, precision, recall, and F1-measure to NB were within 91%, while AUC gave a performance score that was within 95%. Likewise, the performance scores of RF and the proposed model were within the range of 94%, which were given by accuracy, precision, recall, and F1-measure, whereas AUC gave them performance scores that were within 98%. However, SVM was given performance scores close to 93% by all the metrics.

In order to determine whether the proposed model provided a significant improvement in the employee attrition prediction, its performance was compared with that of the baseline comparison algorithms by using a paired t-test, a common statistical test that is used to compare two sets of values [36]. The performance evaluated by AUC was chosen because the AUC metric compares models for the same dataset in terms of their performance; the higher the AUC, the better the performance [30]. Furthermore, it is strongly recommended to use AUC when the dataset is imbalanced. In determining the test base for the paired t-test, RF was chosen because it is a baseline comparison algorithm and has the best performance among the baseline algorithms. The level of significance was set at 0.05. As shown in Table IV, the RF model and the proposed model had an AUC of 98.4% and 98.5%, respectively. According to the paired t-test, the performance of the proposed model was statistically better than that of RF at the significance level of 0.05. Thus, the two-layer nested ensemble model improved the employee attrition prediction.

Table V shows the comparison between the performance of the proposed model in this work and the models in the related work. It is noted that the proposed model achieved great performance. Its performance exceeded the performance of almost all models.

V. CONCLUSION AND FUTURE WORK

Rising employee attrition rates are a major concern that has an influence on an organization's performance [3]. It is affected by the success of the employment process, since the improvement of this process reduces the rate of employee attrition [4]. This paper investigated the effect of using nested ensemble learning technique to improve employee attrition prediction when applied to the *IBM HR Analytics Employee Attrition & Performance* dataset that has been optimized by the information gain-based feature selection method. To the best of the authors' knowledge, no previous research has applied nested ensemble techniques to the *IBM HR Analytics Employee Attrition & Performance* dataset.

The experimental results showed that the performance of RF was better than that of NB and SVM. Despite this, the proposed model outperformed RF. The RF model had an accuracy of 94.2417%, an AUC of 98.4%, a precision of 94.3%, and a recall and an F1-score of 94.2%. However, the proposed two-layer nested ensemble model achieved 94.5255% in accuracy, 98.5% in AUC, and 94.5% in precision, recall, and F1-score. In addition, the proposed model was statistically better than that of RF at the significance level of 0.05. To compare the performance of the proposed model in this work with the models in the related work mentioned in Section II, the proposed model showed excellent performance. Its performance exceeded that of all the models. Accordingly, the findings showed the capability of the proposed model to predict employee attrition. Thus, the nested ensemble learning technique assists the employment process and improve employee attrition prediction.

This paper suggests conducting more studies on the nested ensemble technique as future work. The reviewed studies, such as [37] [22], and [38], which applied nested ensemble techniques, focused on applying an ensemble technique as a meta-level algorithm, whereas in this work, the proposed model is a two-layer nested ensemble model that applied ensemble techniques as base-level algorithms and LR as a meta-level algorithm. However, the proposed model approved its capability in prediction as it scored more than 98% in AUC.

REFERENCES

[1] M. N. Siddiqui, "Success of an Organization is a result of Employees Performance," *Advances in Social Sciences Research Journal*, vol. 1, no. 4, pp. 179–201, Jul. 2014.

[2] S. Garg, S. Sinha, A. K. Kar, and M. Mani, "A review of machine learning applications in human resource management," *Int. J. Product. Perform. Manag.*, vol. ahead-of-print, no. ahead-of-print, pp. 1–21, Feb. 2021.

[3] S. Krishna and S. Sidharth, *Workforce Analytics: Predicting Employee Attrition Using Machine Learning Approach*, 1st ed. London: Routledge, Sep. 2022, pp. 207–221.

[4] S. Gowdru, S. K. Dubli, P. Agarwal, and Bhoomika, "Prediction of Employee Attrition Using Stacked Ensemble Method," in *Information and Communication Technology for Competitive Strategies (ICTCS 2021)*, ser. Lecture Notes in Networks and Systems, M. S. Kaiser, J. Xie, and V. S. Rathore, Eds., vol. 401. Singapore: Springer Nature Singapore, 2023, pp. 451–462.

[5] J. D. MacCharles and E. N. Melton, "Risks of expressing your authentic self in sport: The influence of stigma covering on perceived job fit and hiring recommendations," *SBM*, vol. 12, no. 4, pp. 365–381, Aug. 2022.

[6] F. Guerranti and G. M. Dimitri, "A Comparison of Machine Learning Approaches for Predicting Employee Attrition," *Applied Sciences*, vol. 13, no. 1, p. 267, Dec. 2022.

[7] K. Kanamori, T. Takagi, K. Kobayashi, and Y. Ike, "Counterfactual Explanation Trees: Transparent and Consistent Actionable Recourse with Decision Trees," in *The 25th International Conference on Artificial Intelligence and Statistics (AISTATS) 2022*, ser. Proceedings of Machine Learning Research, vol. 151. Valencia, Spain: Proceedings of Machine Learning Research (PMLR), 2022, pp. 1846–1870.

[8] F. H. Alshammari, "Design of System-of-System Acquisition Analysis Using Machine Learning," *Complexity*, vol. 2022, pp. 1–15, Jul. 2022.

[9] A. Raza, K. Munir, M. Almutairi, F. Younas, and M. M. S. Fareed, "Predicting Employee Attrition Using Machine Learning Approaches," *Applied Sciences*, vol. 12, no. 13, pp. 1–18, Jun. 2022.

[10] S. Porkodi, S. Srihari, and N. Vijayakumar, "Talent management by predicting employee attrition using enhanced weighted forest optimization algorithm with improved random forest classifier," *IJATEE*, vol. 9, no. 90, pp. 563–582, May 2022.

[11] S. Najafi-Zangeneh, N. Shams-Gharneh, A. Arjomandi-Nezhad, and S. Hashemkhani Zolfani, "An Improved Machine Learning-Based Employees Attrition Prediction Framework with Emphasis on Feature Selection," *Mathematics*, vol. 9, no. 11, p. 1226, May 2021.

[12] M. Pratt, M. Boudhane, and S. Cakula, "Employee Attrition Estimation Using Random Forest Algorithm," *BJMC*, vol. 9, no. 1, pp. 49–66, 2021.

[13] N. B. Yahia, J. Hlel, and R. Colomo-Palacios, "From Big Data to Deep Data to Support People Analytics for Employee Attrition Prediction," *IEEE Access*, vol. 9, pp. 60 447–60 458, 2021.

[14] F. Fallucchi, M. Coladangelo, R. Giuliano, and E. William De Luca, "Predicting Employee Attrition Using Machine Learning Techniques," *Computers*, vol. 9, no. 4, pp. 1–17, Nov. 2020.

[15] N. Brockett, C. Clarke, M. Berlingerio, and S. Dutta, "A System for Analysis and Remediation of Attrition," in *2019 IEEE International Conference on Big Data (Big Data)*. Los Angeles, CA, USA: IEEE, Dec. 2019, pp. 2016–2019.

[16] D. K. Srivastava and P. Nair, "Employee Attrition Analysis Using Predictive Techniques," in *Information and Communication Technology for Intelligent Systems (ICTIS 2017) - Volume 1*, ser. Smart Innovation, Systems and Technologies, S. C. Satapathy and A. Joshi, Eds., vol. 83. Cham: Springer International Publishing, 2018, pp. 293–300.

[17] P. Subhash, "IBM HR Analytics Employee Attrition & Performance," https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset, Mar. 2017.

[18] the King Abdulaziz University, Information Systems Department, Faculty of Computing and Information Technology Jeddah, Saudi Arabia, A. Qutub, A. Al-Mehmadi, M. Al-Hssan, R. Aljohani, and H. S. Alghamdi, "Prediction of Employee Attrition Using Machine Learning and Ensemble Methods," *International Journal of Machine Learning and Computing*, vol. 11, no. 2, pp. 110–114, Mar. 2021.

[19] L. Jiang and C. Li, "An Augmented Value Difference Measure," *Pattern Recognition Letters*, vol. 34, no. 10, pp. 1169–1174, Jul. 2013.

[20] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. Waltham, USA: Elsevier/Morgan Kaufmann Publishers, 2012.

[21] A. Burkov, *The Hundred-Page Machine Learning Book*. Polen: Andriy Burkov, 2019.

[22] M. Abdar, M. Zomorodi-Moghadam, X. Zhou, R. Gururajan, X. Tao, P. D. Barua, and R. Gururajan, "A new nested ensemble technique for automated diagnosis of breast cancer," *Pattern Recognition Letters*, vol. 132, pp. 123–131, Apr. 2020.

[23] B. Steele, J. Chandler, and S. Reddy, *Algorithms for Data Science*. Cham: Springer International Publishing, 2016.

[24] M. Ismail, N. Hassan, and S. S. Bafjaish, "Comparative Analysis of Naive Bayesian Techniques in Health-Related for Classification Task," *Journal of Soft Computing and Data Mining*, vol. 1, no. 2, p. 10, 2020.

[25] A. Géron, *Hands-On Machine Learning with Scikit-Learn and TensorFlow_Concepts, Tools, and Techniques to Build Intelligent Systems*, 1st ed. Sebastopol, CA: O'Reilly Media, Mar. 2017.

[26] W. Wang, Y. Li, X. Wang, J. Liu, and X. Zhang, "Detecting Android malicious apps and categorizing benign apps with ensemble of classifiers," *Future Generation Computer Systems*, vol. 78, pp. 987–994, Jan. 2018.

[27] V. Bolón-Canedo and A. Alonso-Betanzos, *Recent Advances in Ensembles for Feature Selection*, ser. Intelligent Systems Reference Library. Cham: Springer International Publishing, 2018, vol. 147.

[28] H. Jiang, *Machine Learning Fundamentals: A Concise Introduction*, 1st ed. Cambridge University Press, Oct. 2021.

[29] M. K. A. Aljero and N. Dimililer, "A Novel Stacked Ensemble for Hate Speech Recognition," *Applied Sciences*, vol. 11, no. 24, p. 11684, Dec. 2021.

[30] A. C. Mu, *Introduction to Machine Learning with Python: A Guide for Data Scientists*, 1st ed. Sebastopol, CA: O'Reilly Media, Inc., 2016.

[31] V. Kotu, *Data Science: Concepts and Practice*, 2nd ed. Cambridge, MA: Elsevier/Morgan Kaufmann Publishers, 2019.

[32] P. Refaeilzadeh, L. Tang, and L. Huan, "Cross-Validation," in *Encyclopedia of Database Systems*, 2nd ed., L. Liu and M. T. Özsu, Eds. New York, NY: Springer New York, Dec. 2018, pp. 389–718.

[33] S. Vluymans, *Dealing with Imbalanced and Weakly Labelled Data in Machine Learning Using Fuzzy and Rough Set Methods*, ser. Studies in Computational Intelligence. Cham: Springer International Publishing, 2019, vol. 807.

[34] K. Das, R. N. Behera, and B. Tech, "A Survey on Machine Learning: Concept, Algorithms and Applications," *IJIRCCE*, vol. 5, no. 2, pp. 1301–1309, Feb. 2017.

[35] W. Dudzik, J. Nalepa, and M. Kawulok, "Evolving data-adaptive support vector machines for binary classification," *Knowledge-Based Systems*, vol. 227, p. 107221, Sep. 2021.

[36] K. Linnet, "Limitations of the Paired t-Test for Evaluation of Method Comparison Data," *Clinical Chemistry*, vol. 45, no. 2, pp. 314–315, Feb. 1999.

[37] M. E. A. Bashir, M. Akasha, D. G. Lee, G. M. Yi, K. H. Ryu, E.-j. Cha, J.-w. Bae, M. C. Cho, and C. W. Yoo, "Nested Ensemble Technique for Excellence Real Time Cardiac Health Monitoring," in *International Conference on Bioinformatics & Computational Biology, BIOCOMP 2010, July 12-15, 2010, Las Vegas Nevada, USA, 2 Volumes*, vol. 2. Las Vegas Nevada, USA: CSREA Press, Jul. 2010, pp. 519–525.

[38] M. Abdar, U. R. Acharya, N. Sarrafzadegan, and V. Makarenkov, "NE-nu-SVC: A New Nested Ensemble Clinical Decision Support System for Effective Diagnosis of Coronary Artery Disease," *IEEE Access*, vol. 7, pp. 167 605–167 620, 2019.