Journal of Media Horizons
ISSN: 2710-4060 2710-4052 Volume 7, Issue 4, 2026

AN EFFICIENT PROCESSİNG-İN-MEMORY COMPUTİNG USING HYBRID CONSENSUS LEARNING BASE-CLASSIFIERS AND CONS MODEL

Fatima Alajaebi[*1], **Dr. Eman Alyasin**[2]

[*1,2]*Information Technology, Altınbaş University, Istanbul, Turkey*

DOI: https://doi.org/10.5281/zenodo.19662072

Keywords
Computer Information Technology (CIT), Normal SQL-traffic or the SQLIA

Article History
Received: 22 February 2026
Accepted: 03 April 2026
Published: 20 April 2026

Corresponding Author: *
Fatima Alajaebi

Abstract
This paper proposes optimization processed for Min-Max normalization that helped alleviating convergence and over-fitting problem. The normalized features were processed for two-class classification using nine machine learning algorithm including Naïve Bayes variants regression techniques, pattern mining, association rule mining, pattern mining, association rule mining, neuro-computing and ensemble learning. These nine base-classifiers constituted a robust heterogenous ensemble learning environment which labelled each SQL-query as the normal traffic or SQLIA and thus based on the maximum voting score our proposed consensus (CONS) model predicted each query as the Normal SQL-traffic or the SQLIA., The resampled features along with the original features were processed for feature selection using significant predictor test and variance threshold feature selection (VTFS) algorithms so as to improve time-efficiency as well as to reduce redundant computation. The selected feature.

I. INTRODUCTION

In this paper a state-of-art new and robust NLP semantic feature driven heterogenous consensus model based SQLIA prediction model is proposed for SCPM purposes (NSF-HC SQLIAS) is proposed. Unlike classical approaches, our proposed NSF-HC SQLIAS model employs semantic feature extraction followed by data-sampling and feature selection to refine set of suitable features for final classification. The proposed NSF-HC SQLIAS model at first applies tokenization and stopping [1] word removal that converts each query into the set of tokens, while stopping words removal reduces redundant terms or artifacts to make data more suitable. Subsequently, it performs lower case transformation in which each word is converted into lower case. Once tokenizing each SQL query, NSF-HC SQLIAS applies the well-known semantic feature extraction methods named Word2Vec and TF-IDF to extract the feature. Noticeably, there has been argument that amongst the word2vec variants, CBOW and SKG have the contradictory performance towards NLP problem. Considering this fact, NSF-HC SQLIAS applies both CBOW as well as SKG along with TF-IDF as distinct feature extraction method[2.3]. Once extracting the features, realizing data imbalance or class-imbalance problem, we performed re-sampling using random sampling and up-sampling methods. The resampled features were processed for feature selection, to reduce redundant computation. Specifically, NSF-HC SQLIAS applied two distinct algorithms Wilcoxon rank sum test and variance threshold estimation method to select most efficient set of features for further model training and classification. The selected features were later processed for Min-Max normalization that helped avoiding convergence, local minima and over-fitting problems. Finally, to improve SQLIA prediction accuracy and reliability[4,5], a state-of-art heterogenous ensemble model was designed to provide consensus-based classification. The

proposed consensus-based learning applied nine classifiers including Naïve Bayes (MNB) classifier variants (i.e., Multinomial, Bernoulli and Gaussian), Support Vector Machine with radial basis functions (SVM-RBF), Logistic Regression (LOGR), Decision Tree (DT), Artificial Neural Network with Levenberg Marquardt (ANN-LM), Random Forest (RF) and AdaBoost (ADAB) algorithms. These algorithms were applied distinctly to perform two-class classification and hence classified each query as Normal Traffic or SQLIA, and labeled it with "0" and "1", respectively. Thus, with the maximum voting, the proposed consensus model predicted each SQL-query as Normal or SQLIA. To identify the set of optimal performing environment, NSF-HC SQLIAS was examined for the different feature [7,8] extraction methods, resampling, feature selection methods as well as the varied standalone classifiers. The performance characterization was achieved in terms of SQLIA prediction accuracy, F-Measure and AUC.

2. Problem Formulation and Hypothesis

This research primarily intends to explore whether the use of a robust machine learning model with large semantic feature training can yield optimal SQLIA prediction for IoT-driven SCPM systems. Being a machine learning model, to achieve above stated goals, the optimization has been incorporated at the different levels of the computing environment, i.e., data pre-processing, feature extraction, selection and classification. Predominantly, it intends to design (or eventually achieve) a set of robust computing environment which could perform optimal SQLIA prediction without undergoing convergence, local minima, overfitting and false-positive prediction. To achieve it, optimizing both data model as well as computing environment is must. In this reference, towards data model enhancement, the proposed model hypothesizes that unlike classical structural or topological information learning based approaches, the use of semantic features can yield higher level of ontological feature-learning and hence can help improving SQLIA prediction accuracy, even under camouflaged injection efforts or mimicking efforts. To achieve it, the use of semantic feature extraction models such as Word2Vec word-embedding can be vital. Though, there are the different variants of word-embedding, the performance assessment of CBOW, SKG and TF-IDF can help identifying the best performing feature-set. Moreover, in addition to the semantic feature learning addressing class-imbalance problem is inevitable, which has not been addressed so far by any researcher, especially amongst the researches towards SQLIA prediction. To alleviate this problem resampling seems to be a viable solution[9,10]; however, their efficacy towards SQLIA-prediction is yet to be examined. In this reference, this research employs multiple re-sampling methods including random sampling, up-sampling over the extracted semantic features (i.e., CBOW, SKG and TF-IDF features) and examines their respective performance to identify the best suitable sampling method and allied feature. In sync with this goal, we hypothesize that the use of sampling technique(s) over semantic features can achieve higher reliability by avoiding class-imbalance problem. This is the matter of fact that the use of resampling can introduce a set of additional feature pattern or examples to make learning more efficient; however, at the cost of increased computation and time, which cannot be suggested for real-time IoT-driven SCPM purposes. In the proposed work, feature selection method is proposed. Here, feature extraction methods including significant prediction test and variance threshold methods are applied over the resampled features. Consequently, it helps retaining only those features having impact on eventual SQLIA prediction results. Furthermore, to alleviate the over-fitting problem, the proposed model employs Min-Max normalization that can map entire data instances in the range of 0-1 and hence can help avoiding pre-mature convergence as well. Once normalizing the feature vector(s), it can be processed for two-class classification. Though, a few efforts have been made by using machine learning algorithms for SQLIA prediction; however, the diversity of performance by the different standalone algorithms puts question mark on the acceptability of any single solution. For instance, the different machine learning algorithms with the same input features give different accuracy or allied

performance signifying diversity of performance. In this case, applauding a single model with single highest performance seems unjustifiable. Considering this fact, in this research a heterogenous ensemble learning model is developed using machine learning classifiers from the different categories such as pattern learning, regression, neuro-computing and ensemble. Here, the key motive is to exploit performance by each base-classifiers to make an eventual consensus to classify an SQL query as Normal or SQLIA. To achieve it, the proposed model employed a total of nine machine learning algorithms including MNB, GNB, BNB, SVM-RBF, LOGR, DT, ANN-LM, RF and ADAB algorithms. Functionally, these algorithms classify each SQL query as Normal or SQLIA and labels as "0" and "1", correspondingly. In this manner, the proposed Consensus based model identifies the maximum score for each query and hence a query with higher (here, 5 score) 1's is predicted as SQLIA query while higher 0's indicates the normal traffic. In this manner, the proposed model classifies each SQL query as Normal or SQLIA. The key benefit of this approach can be higher reliability due to consensus-based prediction in comparison to the standalone classifiers. In this reference, it hypothesizes that the inclusion of consensus-based learning with heterogenous ensemble (learning) paradigm can achieve better performance and reliability. Since, the overall proposed model follows the natural language processing concept, the overall process can be defined as an NLP problem. In this reference, the overall contribution of this research can be introduced as "NLP-Driven Semantic Featured Heterogenous Consensus Learning Model for SQLIA Prediction for IoT-Smart City Infrastructures (NSF-HC SQLIAS).

3 RESULT AND DISCUSSION

The above table (Table I) illustrates some of the examples of the normal and SQLIA traffics. The corresponding class labels (i.e., 0 for the normal traffic and 1 for SQLIA) is given in Table I. Most of the existing methods as discussed in Section II, employs structural learning approach to classify queries; however, learning semantic ontological information can be relatively better to perform SQLIA [11][12]. Now, recalling the at-hand SQLIA problem as NLP-problem, we performed automated approach to pre-process the input queries and convert it into the uniform set of tokens for further feature extraction. In the proposed model, at first tokenization was performed over the input queries. The proposed tokenization model converted each query into a set of tokens which were later used for word-embedding and allied feature extraction. Observing Table I, it can be found that the queries contain stopping words and punctuation marks and hence removing such redundant components is vital. To achieve it, we performed stopping words and punctuation mark removal followed by lower case conversion. In the proposed model, each token was converted into lower-case. Thus, the above stated tokenization and allied pre-processing methods generated a set of tokens signifying the frequently occurring words. It transformed input query corpus into different tokens. Now, once creating the set of tokens from each query, our proposed NSF-HC SQLIAS model executed feature extraction method.

The classification results with the different standalone classifiers as well as CONS model are given in Fig. 1.2.3 shows the accuracy performance by the different base classifiers and CONS model. As depicted in the simulation results, the proposed CONS model outperforms other methods with the highest accuracy of 98.4%. Interestingly, in addition to the proposed maximum voting ensemble assisted CONS model, LOGR too has exhibited the accuracy of 98.2%, followed by ANN-LM (98.0%), SVM-RBF (97%), ADAB (95.7%), GNB (92.1%), DT (90%), MNB (90%), RF (86%) and BNB (85%).

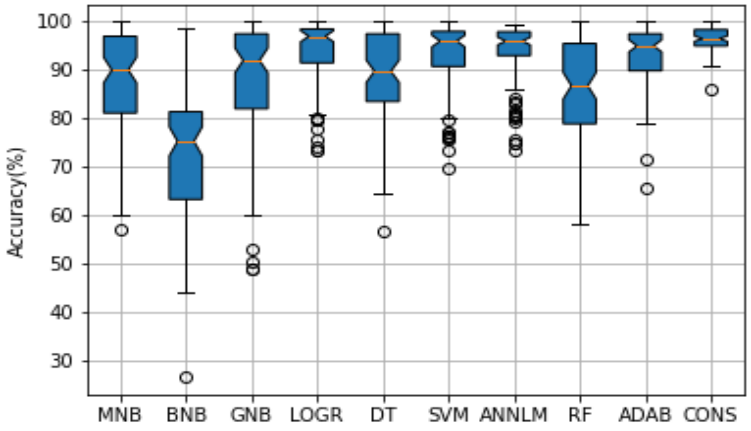

Fig. 1 Accuracy with the different base-classifiers and CONS model

Noticeably, unlike base-classifiers the proposed CONS model his outperformed other base-classifiers in terms of higher accuracy and reliability. It affirms suitability of the proposed model towards SQLIA prediction. To be noted these results (Fig. 1 to Fig. 3) have been obtained with CBOW features followed by random sampling and VTFS feature selection (with normalized feature). Fig. 1 shows F-Measure performance by the different classification methods. As depicted in Fig. 2, the proposed CONS model exhibits the highest F-Measure of 0.993, followed by ANN-LM (0.974), ADAB (0.972), LOGR (0.971), SVM-RBF (0.970), GNB (0.918), MNB (0.900), DT (0.892), RF (0.873), and BNB (0.762).

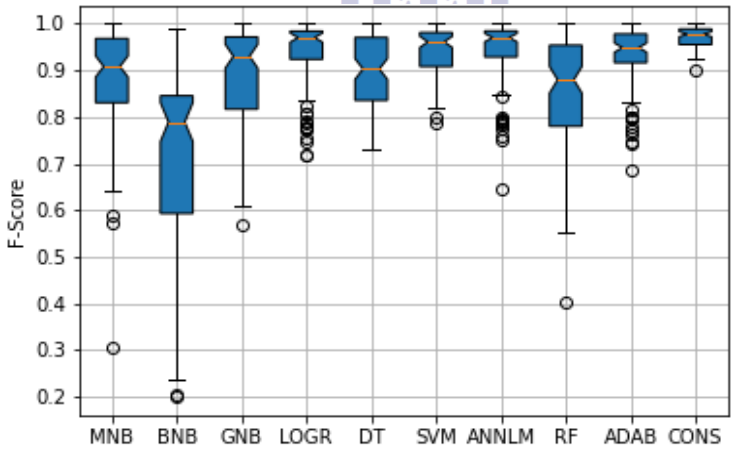

Fig. 2 F-Measure with the different base-classifiers and CONS model

Noticeably, despite of the higher accuracy, the outliers with ANN-LM and ADAB are more signifying corresponding F-score below 0.9. With depth assessment, we find that CONS model is better towards SQLIA prediction. Fig. 13 shows the AUC performance by the different base-classifiers and eventual CONS model. As depicted through the results (Fig. 3), it can be found that the AUC performance with the proposed consensus model (CONS) is superior (AUC=0.99) than other base classifiers. Though, other machine learning methods like ANN-LM (0.989), LOGR (0.99), SVM-RBF (0.98), ADAB (0.976) too has exhibited better AUC, confirming their suitability

towards SQLIA; however, the robustness and efficacy of CONS confirms its optimal utilization for at hand SQLIA prediction and NSF-HC SQLIAS goal(s).

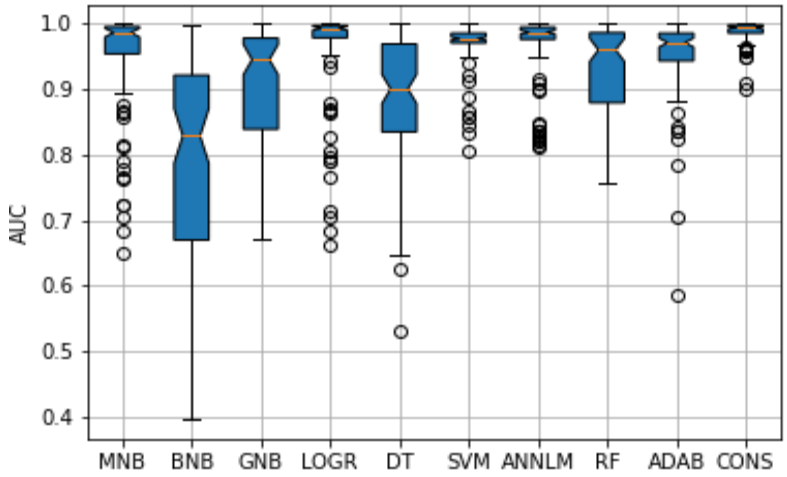

Fig. 3 AUC with the different base-classifiers and CONS model

Observing performance, it can easily be found that the use of CBOW semantic features in conjunction with random sampling, VTFS feature selection and Min-Max normalization, followed by the heterogenous ensemble learning model can yield optimal performance towards NSF-HC SQLIAS. Therefore, segmenting out the best performance by our proposed model (Table II), we performed inter-model performance characterization.

In general, SQL-query's tokens encompass different keywords operators, literals, identifiers, and punctuation marks. Additionally, the SQL syntax language component embodies SQL clause, predicate and expression. Here, learning syntactic traits and allied latent information can help classifying the query as the normal or SQLIA. In a typical SQL-query, clause, predicate and expressions are often employed to control data structure that helps identifying the presence of a malicious data element or allied sot per query. On the other hand, intruders or attackers often use aforesaid design constructs and allied spaces to intrude intrusion. Typically, attackers use such spots or locations to intrude inside the database. Similarly, such spots can be considered for SQLIA detection decisions. This is the matter of fact that the majority of the existing systems employ identification of such standard structural elements or spot to perform SQLIA prediction; however, are limited due to camouflage or mimicking-based attackers or bots. In this case, the use of semantic features amongst the tokens can be viable solution towards cognitive decision for SQLIA. To achieve it, in this paper our proposed NSF-HC SQLIAS model employed three different semantic

3. DİSCUSSİON

1. Accuracy is "the degree to which a measurement's result corresponds to the appropriate value or a standard," and it relates to how closely a statistic is to its acceptance rate. *Heterogenous Ensemble assisted Consensus Model for SQLIA Prediction*

In NSF-HC SQLIAS model, nine different base classifiers belong to association rule mining, regression, pattern mining, neuro-computing and ensemble methods taken into consideration to design a heterogenous ensemble learning environment. Here, our key motive was to exploit diversity of performance by the different standalone machine learning models to obtain consensus (CONS) so as to make final classification decision. To achieve it, NSF-HC SQLIAS applied nine different machine learning models comprising GN, BNB, MNB, LOGR, DT,

SVM-RBF, ANN-LM and ADAB classifiers. The aforesaid classifiers were applied as standalone machine learning model to classify each SQL query as the normal traffic or the SQLIA query. Additionally, these classifiers labelled each query as "1" and "0" for the normal traffic and SQLIA-query, respectively. To perform CONS based prediction, the labels by all base classifiers (i.e., nine algorithms as mentioned above) were employed to estimate maximum voting and the query with minimum five 0's was predicted as the normal traffic while the one with minimum 1's was classifiers as SQLIA query.

Table 1 Concluding performance by the proposed NSF-HC SQLIAS model

Algorithm	Accuracy (%)	F-Measure	AUC
NSF-HC SQLIAS	**98.4**	**0.993**	**0.999**

Inter-Model Performance Characterization

In order to examine the relative performance by our proposed model, we performed inter-model characterization by comparing accuracy, F-Measure and AUC metrics of the proposed NSF-HC SQLIAS model as well as other.

5. CONCLUSION

İn this research focused on employing different set of feature extraction, selection and sampling methods, the relative performance revealed that undeniably semantic features show better performance; however, amongst TF-IDF. SKG and CBOW, TF-IDF and CBOW performed superior with more than 98% of accuracy, F-Measure of more than 0.98 and similarly AUC of 0.970-0.999. It shows robustness of the semantic features towards SQLIA prediction task. Interestingly, the results from sampling method were contradictory where even original data exhibited near similar performance like the random sampling-based features. However, realizing undeniable probability of class-imbalance in realistic SQL traffic, the use of random sampling seems more viable solution. Moreover, the use of correlation based significant predictor test exhibited satisfactory performance; though the efficacy of VTFS was competitive to it. It indicates that the use of either of these methods can be vital. It can retain better efficiency even with reduced sample size and hence computation. Finally, unlike classical standalone classifier or machine learning the proposed maximum voting consensus (CONS) based model exhibited superior performance than any of the standalone classifier. Interestingly, being a consensus-based classification, the reliability of the proposed NSF-HC SQLIAS model is higher and justifiable. The highest accuracy of NSF-HC SQLIAS model was found 98.6%, F-Measure of 0.993 and AUC of 0.999, which can be called satisfactory and applicable towards IoT-driven SCPM infrastructure security. In future, some other feature extraction, selection methods along with deep learning concept(s) can be applied to assess relative efficacy for SQLIA prediction in SCPM infrastructures.

REFERENCE

[1] Narzieva, D.M. and Kudratov, A.D., 2021. the importance of digitalisation of the economy and priorities in Uzbekistan. World Economics and Finance Bulletin, 2(2), pp.9-13.

[2] Mamatzhonovich, O.D., Khamidovich, O.M. and Esonali o'g'li, M.Y., 2022. digital economy: essence, features and stages of development. Academicia Globe: Inderscience Research, 3(04), pp.355-359.

[3] Glushchenko, T.E., Khodarinova, N.V., Ishchenko, O.V., Shaposhnikov, V.L. and Aksenova, Z.A., 2021. The development of cooperation in the digital economy based on scientific research by av chayanov. In Frontier Information Technology and Systems Research in Cooperative Economics (pp. 43-52). Springer, Cham.

[4] Lu, C.W., Huang, J.C., Chen, C., Shu, M.H., Hsu, C.W. and Bapu, B.T., 2021. An energy-efficient smart city for sustainable green tourism industry. Sustainable Energy Technologies and Assessments, 47, p.101494.

[5] Sheng, J., Amankwah-Amoah, J., Khan, Z. and Wang, X., 2021. COVID-19 pandemic in the new era of big data analytics: Methodological innovations and future research directions. British Journal of Management, 32(4), pp.1164-1183.

[6] Kobilov, A.U., Khashimova, D.P., Mannanova, S.G. and Abdulakhatov, M.M.O., 2022. Modern Content and Concept of Digital Economy. International Journal of Multicultural and Multireligious Understanding, 9(2), pp.375-378.

[7] Zolkin, A.L., Shavanov, M.V. and Hmizova, N.G., 2021, March. Features of the use of information technologies in agriculture. In IOP Conference Series: Earth and Environmental Science (Vol. 677, No. 3, p. 032091). IOP Publishing.

[8] Hakimovna, A.F. and Bebudovna, B.S., 2021. Wide dissemination of information technologies in the sphere of domestic and international tourism. Economics, (1 (48)), pp.32-34.

[9] Gibadullin, A., Digital and Information Technologies in Economics and Management.

[10] Valijonovna, K.I. and Askaraliyevna, U.M., 2021. Use of Modern Information Technologies in Literacy Classes. International Journal of Culture and Modernity, 11, pp.268-273.

[11] Zolkin, A.L., Domracheva, E.A., Losev, A.N. and Avdeev, Y.M., 2021, February. Use of the modern information technologies for continuous monitoring of transport infrastructure. In IOP Conference Series: Materials Science and Engineering (Vol. 1047, No. 1, p. 012094). IOP Publishing.

[12] Su, Y. and Wang, X., 2021. Innovation of agricultural economic management in the process of constructing smart agriculture by big data. Sustainable Computing: Informatics and Systems, 31, p.100579.