Mapping Graduate Skills to Market Demands: A Holistic Examination of Curriculum Development and Employment Trends

Abdulsamad Ebrahim Yahya
Department of Information Technology, Faculty of Computing and Information Technology, Northern Border University, Rafha 91911, Saudi Arabia
abdulsamad.qasem@nbu.edu.sa

Wael M. S. Yafooz
Computer Science Department, College Computer Science and Engineering, Taibah University, Medina, 42353, Saudi Arabia
wyafooz@taibahu.edu.sa (corresponding author)

Atef Gharbi
Department of Information Systems, Faculty of Computing and Information Technology, Northern Border University, Rafha 91911, Saudi Arabia
atef.gharbi@nbu.edu.sa

Received: 10 April 2024 | Revised: 30 April 2024 and 5 May 2024 | Accepted: 6 May 2024

ABSTRACT

The number of unemployed computer science graduates has increased significantly over the last few years. The primary reason for this problem is the skill gap between the graduates and what is required on the job market. The current study aims to address the challenge of aligning the skills of computer science graduates with the evolving demands of the job market. To achieve this objective, the current research leverages Machine Learning (ML) and Deep Learning (DL) techniques to predict the skills required by employers and those possessed by graduates. The dataset used in this study has been carefully curated and annotated by experts in the field. It entails 18 features that capture various aspects of a graduate's skillset, such as programming languages, technical expertise, and soft skills. Additionally, the dataset includes information on the most common job positions in the computer science industry (i.e. a total of 8 roles). A sample size of 3,831 computer science graduates was sourced from alumina surveys and reputable hiring agencies. The dataset provides a comprehensive view of the skills landscape in the computer science domain. Several ML classifiers, ensemble methods, and DL approaches were utilized in a series of experiments. The correlations and important skills and jobs in the market were given focus. The experimental results indicate that support vector machines and neural networks achieved high accuracies of 82% and 88%, respectively. By analyzing the results, this study seeks to uncover patterns and trends that can guide the development of educational programs and curricula, ensuring they are aligned with the evolving needs of the industry.

Keywords-computer science graduates; machine learning; deep learning; job market; academic programs

I. INTRODUCTION

During the recent years the computing field has witnessed exponential growth, with technologies, such as artificial intelligence, cloud computing, and data science becoming integral components of industries worldwide. However, this rapid evolution has created a significant gap between the skills possessed by students graduating from computing schools and the needs of the job market [1-3]. This disparity is primarily caused by the traditional curriculum's inability to keep pace with the dynamic nature of the industry, resulting in graduates lacking the essential skills required by employers. The market demands computer science professionals who possess strong technical skills and also demonstrate proficiency in problem-solving, critical thinking, and collaboration. Employers seek individuals who can immediately adapt to new technologies, communicate effectively, and work efficiently in multidisciplinary teams. Unfortunately, many computing programs fail to adequately address these needs, leading to a mismatch between the skills students acquire and those demanded by the industry.

To bridge the aforementioned gap, educational institutions must revamp their computing curriculum, particularly by incorporating hands-on projects, real-world case studies, and industry partnerships [4]. By doing so, students can develop the necessary skills and competencies required to meet the demands of the ever-evolving computing industry. Additionally, fostering a culture of lifelong learning and professional development amongst students can help them stay abreast of the latest trends and technologies, ensuring they remain valuable assets to prospective employers. Researchers and educators have recognized the pressing need to address the discrepancy between students' skills in the computer science field and job market requirements [5-8]. An effort has been exerted to improve course learning outcomes and quality procedures to better equip students with the skills demanded by employers. However, challenges remain in ensuring that graduates possess the necessary skills to excel in the rapidly evolving computing industry. To tackle these challenges, researchers have turned to Machine Learning (ML) [9-13], Natural Language Processing (NLP) [12, 13-15], and Deep Learning (DL) approaches [16-18] in various areas, including education. These technologies have shown promise in enhancing learning outcomes, personalizing the education and bridging the gap between academia and industry needs.

This study proposes and evaluates the efficacy of ML classifiers and Artificial Neural Network (ANN) models in categorizing graduate skills to match the demands of the job market. To achieve this goal, this study introduces a novel dataset comprising 18 features, representing the requisite skills for 8 prominent computer science jobs. This dataset was collated from three universities in Saudi Arabia and Malaysia and supplemented by job descriptions from hiring agencies. Additionally, three experts with PhDs and administrative experience in computer science departments validated the dataset. By applying sophisticated ML models to the suggested dataset, this study's goal is to contribute to narrowing the gap between computer science students and the skill sets sought by employers. Several experiments have been conducted, and the results disclosed that ANNs and Support Vector Machine (SVM) classifier recorded the best accuracy of 88% and 82%, respectively. The main contributions of this study are:

- ML classifiers, ensemble voting, and ANN models are trained to classify the required skills with their corresponding jobs in computer science, thereby advancing classification techniques for job–skill alignment.
- The most required skills are identified through correlation matrices and feature importance analysis.
- A novel dataset is introduced, comprising 3,831 rows of 18 required skills and 8 demanded jobs in the computer science field. The dataset serves as a valuable resource for further research and provides a comprehensive snapshot of the current market demands in the computer science sector.

II. RELATED STUDIES

This section provides an overview of the related studies focusing on the employment outcomes of graduate students, the demand for specific skills in the job market, and the development of curriculum to meet industry needs. Authors in [12] predicted the dropout rate of nontraditional undergraduate students with a balanced accuracy of 78.44% and an F1 score of 73.76%. The XGBoost ML version, which was employed, beat traditional tactics in factors like financial independence, caregiving responsibilities, and enrollment styles using data from the Beginning Postsecondary Students Longitudinal Study. The study emphasized the importance of factors, such as total mortgage amounts, Pell Grants, and college students' projected chance of degree final touch. Authors in [10] considered the application of a computational statistics software and system learning techniques to improve decision-making regarding student employment. The utilization of sophisticated algorithms to look for information trends, predict placement outcomes, and identify factors affecting pay and placement notoriety was emphasized. Authors in [11] identified essential technical skills for AI and ML roles by examining active listings. They found that compared to AI positions, ML positions placed greater focus on useful technical skills. To determine the necessary skills, the writers conducted a content analysis of Indeed.Com job adverts. Authors in [9] discussed the use of ML models in software to categorize online task classified advertising for tracking the hard labor market in real time. Authors in [19] investigated the disparity between the capabilities held by business graduates and the competencies required by managers in the banking sector. The results highlight how important it is for academic institutions to build stronger relationships with business in order to better satisfy industry needs and increase graduates' employability. Authors in [20] conducted a thorough gap analysis utilizing job posting data from the Emsi database. Data Science, Automation, Cyber, and Sensors are the four technology clusters that the authors identified based on their labeling of skills and domain knowledge in accordance with Industry 4.0 pillars. The survey found large gaps in the labor force, with employer demand exceeding worker supply for skills like Python, big data tools (Hadoop, Spark), cloud platforms (AWS, Azure), and domains including device learning, cybersecurity, and algorithms. Proficiency in programming languages, such as Python, SQL, C++, and Java became highly desirable across several clusters. Authors in [21] compared the opinions of Slovakian employers and college students on the value and satisfaction of university graduates' employability skills. Students placed more weight on topic experience, leadership, and subject knowledge, whereas employers emphasized abilities, such as engagement, duty, and adaptability. It was found that the two groups' levels of pride and ability significance differed significantly. The analysis also emphasized how difficult it is to find recent graduates and how important it is to match training objectives with the business goals. In [22], a course and skill recommender system was developed to provide students and learners with personalized course and skill recommendations that will help them enhance their position in the software job market.

III. PROBLEM FORMULATION

The problem of this study is defined as a multi-class classification problem. Let X be the set of features representing the skills of computer science graduates and Y be the set of job positions in the market. Each instance x_i in X is a feature vector representing the skills of a computer science graduate, and each corresponding label y_i in Y is the job position that the graduate is seeking. The goal is to learn a mapping f:X→Y that predicts the job position y given the skills x of a computer science graduate. A loss function L over the training dataset $D=\{(x_1,y_1),(x_2,y_2),...,(x_n,y_n)\}$ can be minimized by:

$$f^* = \arg\min_f \frac{1}{n}\sum_{i=1}^{n}(f(x_i), y_i) \quad (1)$$

where f^* is the optimal function, and L is a loss function that measures the difference between the predicted job position $f(x_i)$ and the actual job position y_i. Cross-entropy loss and hinge loss are common loss functions for classification tasks.

The function f can be implemented using various machine learning algorithms, such as SVMs, Decision Trees (DTs), Logistic Regression (LR), Random Forest (RF), and Gradient Boosting (GB). The performance of the model can be evaluated deploying metrics, such as accuracy, precision, recall, and F1-score on a separate test dataset.

IV. MATERIALS AND METHODS

This section outlines the methods and phases utilized to achieve the primary objective of the current study: predicting the skills required in the job market and matching computer science graduates to these skills. The process begins with the establishment of a dataset, involving the extraction of skills from program learning outcomes and alumni surveys. Subsequently, an initial dataset is constructed, followed by a preprocessing phase to prepare the data for analysis. Thereafter, ML and DL models are applied to the preprocessed data, and the performance of these models is evaluated. These phases are presented in Figure 1.

A. Data Collection

Data were collected from graduate students and academic staff via an alumni survey and from from job descriptions from several hiring companies. Before designing the survey, graduates' skills were identified based on the course and program learning outcomes of five programs: Bachelor's degree in Computer Science, Bachelor's degree in Information Technology, Bachelor's degree in Computer Engineering, Bachelor's degree in Information Systems, and Bachelor's degree in Software Engineering. Three experts with PhDs in computer science, who held leadership positions (e.g. heads of departments and college deans), contributed by identifying the most common skills amongst the graduate students. Table I presents the final list of the required skills and Table II portrays the job demands available in the market.

B. Data Preprocessing

The data went through several preprocessing steps to ensure quality and suitability for further analysis. Firstly, data cleaning was performed to remove duplicate entries and irrelevant data and to handle any missing values. Secondly, categorical variables, such as types of skills, were encoded into numerical values to make them usable for analysis. Datasets from graduates and job descriptions were integrated, and the identified skills from program learning outcomes and expert input were also included.

C. ML and DL Models

This section provides an overview of the ML and DL models employed in this study. The models are grouped into three categories to address the objective of predicting and matching skills, which are required on the job market. The first category is classical ML classifiers, involving SVM, DTs, LR, RF, Stochastic Gradient Descent (SGD), GB, AdaBoost, and Naive Bayes (NB) [23, 24]. These models form the framework for skill prediction and matching.

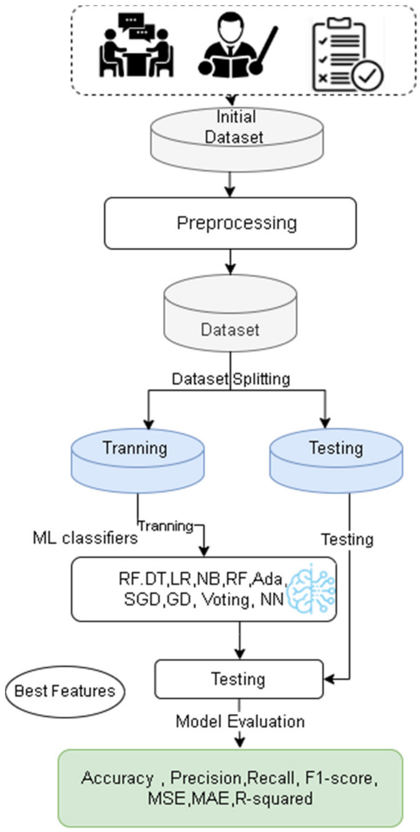

Fig. 1. Research methodology.

TABLE I. LIST OF REQUIRED SKILLS

No.	Required skills	No.	Required skills
1	Project management	10	C++
2	Python	11	SQL
3	Java	12	UML
4	Deep learning	13	Statistics
5	Machine Learning	14	Git
6	CSS	15	NLP
7	HTML	16	Problem solving
8	Security	17	Communication
9	Networking	18	Team work

TABLE II. LIST OF JOB POSITIONS

No.	Job Position	No.	Job Position
1	Software Developer	5	AI Researcher
2	Network Engineer	6	Machine learning
3	Systems Analyst	7	Data Scientist
4	Web Developer	8	Cybersecurity analyst

- LR is able to solve binary classification problems with the help of a simple linear model. Based on a threshold, it estimates probabilities and predicts outcomes.
- SGD is often used for large-scale ML tasks. Whenever a training example's gradient changes, the model parameters are updated.
- SVM is an ML algorithm that finds the best way to separate different classes in feature space. A high-dimensional space and cases with more dimensions than samples make it effective.
- NB is a probability classifier based on Bayes' theorem, but with strong (naive) assumptions about the independence of features.
- DT represents features as internal nodes, decisions as branches, and outcomes as leaf nodes. Both classification and regression tasks can be handled by this algorithm.
- In RF, DTs are grouped to form a predictive model. The algorithm is known for its high accuracy and ability to handle large datasets with many features.
- In AdaBoost, multiple weak classifiers are combined to create a stronger classifier. Correctly classified instances are iteratively adjusted to focus on difficult cases.
- GB is a method that also combines weak learners into one powerful learner. Nevertheless, it fits the newly developed model to residual errors made by the previous model in a different way than AdaBoost.

The second category involves a voting method [25] combining the predictions of the three best classifiers—SVM, LR, and RF—based on their individual performance in terms of accuracy. This ensemble approach aims to enhance prediction robustness and accuracy. The third category focuses on ANNs, representing the DL aspect of the study. ANNs offer a complex, nonlinear approach to skill prediction, enabling the exploration of intricate patterns and relationships within the dataset. By utilizing this diverse range of ML and DL models, this study aims to analyze skill prediction and matching, ultimately presenting the development of educational programs and curricula tailored to meet industry demands.

D. Model Evaluation

In assessing the performance of the models, a variety of widely used metrics were deployed [26, 27]. These metrics are crucial in terms of quantifying the effectiveness of the models in predicting and matching job market skills.

Accuracy measures the proportion of correctly predicted matches between graduates' skills and job requirements:

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \quad (1)$$

Precision assesses the models' accuracy in identifying relevant skills for a given job and is calculated by:

$$\text{Precision} = \frac{TP}{TP+FP}. \quad (2)$$

Recall evaluates the models' ability to capture all relevant skills needed for a particular position and is calculated by:

$$\text{Recall} = \frac{TP}{TP+FN}. \quad (3)$$

F1-score is the harmonic mean of precision and recall. It offers a balanced assessment, ensuring that the precision and recall aspects are adequately considered in evaluating the models' effectiveness.

$$\text{F1}-\text{Score} = 2\,\frac{Recall\ X\ Precision}{Recall+Precision} \quad (4)$$

The Mean Squared Error (MSE) is also used to assess the accuracy of a prediction model. MSE is calculated as the average of the squared differences between the predicted and the actual values:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (5)$$

where n is the total number of skills, y is the actual and $\hat{y}$ is the predicted value.

Mean Absolute Error (MAE) is calculated as the average absolute difference between the predicted and the actual values:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (6)$$

R^2 is a measure to assess the performance of the model and is represented by:

$$R^2 = 1 - \frac{SS_{residual}}{SS_{total}} \quad (7)$$

V. RESULTS AND DISCUSSION

A. Experiment Setting

The experiments were conducted utilizing Jupyter with Python 3.6.1, which is an open-source web application for executing ML and DL experiments. These experiments were conducted utilizing four main packages. Sklearn was applied to split datasets, extract features, build ML classifiers, and evaluate confusion matrices and models. NLTK was engaged for tokenization and stopword removal. Matplotlib was employed for plotting accuracy and validation figures and the confusion matrix. The seaborn package was utilized for displaying the correlation and importance figure in a heatmap. Finally, the DL model was implemented using the TensorFlow framework. The scikit-learn library was deployed for implementing the ML classifiers SVM, DT, LR, RF, and GB. For the DL models, sequential models were put into service using frameworks, such as TensorFlow or Keras. The dataset, consisting of 18 features and information from 3,831 computer science graduates, was preprocessed and divided into training and testing sets utilizing an 80:20 ratio. The parameters of the ML classifiers are presented in Table III and those for the ANN are displayed in Table IV. The most common measurements deployed are precision, recall, accuracy, F1-score, MAE, MSE, and R^2.

B. Results and Discussion

Table V presents the performance metrics of the considered ML classifiers. The voting classifier combines the predictions of SVM, LR, and RF models.

Table V manifests the experimental results and shows the models' performance in terms of accuracy, precision, recall and F1-score. GB achieved an accuracy of 83.36%, with precision, recall, and F1-score above 82%, indicating a strong overall performance. RF also performed well, with an accuracy of 82.92% and similar scores for precision, recall and F1-score. LR, SVM, and DT had accuracies above 81%, whilst Ada, NB, and SGD classifier had lower accuracy scores. RF, GB, LR, SVM, DT, and voting classifiers all had precision scores above 81%, indicating their ability to avoid false positives. NB had the highest precision at 64.99% but was accompanied by lower recall and F1-score.

TABLE III. PARAMETERS OF ML CLASSIFIERS

Classifier	Parameters
SVM	kernel='rbf', C=1.0, gamma='scale', decision_function_shape='ovr'
DT	criterion='gini', splitter='best', max_depth=None
LR	penalty='l2', C=1.0, solver='lbfgs', max_iter=100
RF	n_estimators=100, criterion='gini', max_depth=None
GB	n_estimators=100, learning_rate=0.1, max_depth=3
Ada	base_estimator=None, n_estimators=50, learning_rate=1.0, algorithm='SAMME.R', random_state=None
NB	priors=None, var_smoothing=1e-9
SGD	loss='hinge', penalty='l2', alpha=0.0001, l1_ratio=0.15, fit_intercept=True, max_iter=1000, tol=0.001

TABLE IV. ANN HYPER PARAMETERS

Hyperparameter	Value
Optimizer	Adam
Learning rate	0.0001
Loss function	Sparse categorical cross entropy
Number of epochs	50
Batch size	32
Dropout rate	0.5
Number of hidden layers	2
Number of neurons	128
Activation function	ReLU (Hidden Layers), Softmax (Output Layer)

TABLE V. ML CLASSIFIER COMPARISON

ML classifier	Accuracy	Precision	Recall	F1-score
RF	82.92%	83.01%	82.92%	82.49%
Ada	48.03%	31.40%	48.03%	35.77%
GB	**83.36%**	83.32%	83.36%	82.82%
LR	81.31%	81.39%	81.31%	80.50%
SGD	79.71%	79.66%	79.71%	79.20%
SVM	81.46%	81.41%	81.46%	80.33%
NB	54.16%	64.99%	54.16%	47.73%
DT	81.61%	81.51%	81.61%	81.24%
Voting	81.75%	81.50%	81.75%	80.88%

Figure 2 demonstrates a comparative analysis of the classifiers' performance in predicting the required skills for computer science jobs. MAE measures the average magnitude of the errors in predictions, with lower values indicating better accuracy. MSE quantifies the average squared difference between the predicted and the actual values, with lower values denoting closer prediction to the actual values. An R^2 value near 1 suggests a better fit between the model and the data. The results signal that GB and the voting classifier performed relatively well, with low MAE and MSE and high R^2 values, indicating their effectiveness in predicting required skills for computer science jobs. The confusion matrix of GB and voting classifiers can be seen in Figures 2 and 3, respectively. The correlation between the required skills and the available jobs in the market for computer science graduates is presented in Figure 5. The features' importance is observed in Figure 6.

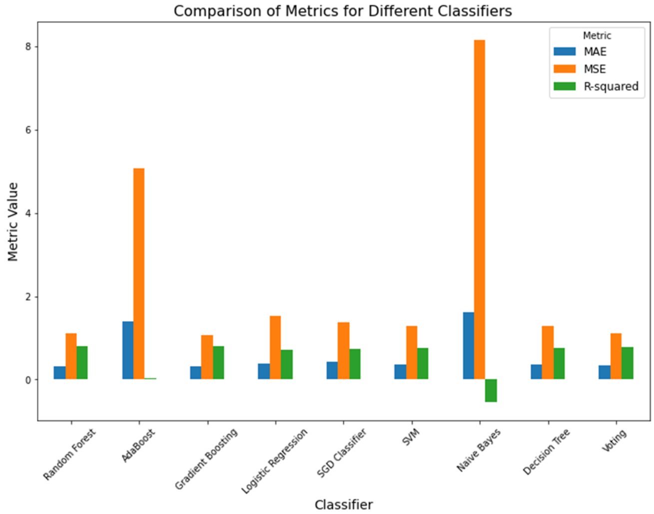

Fig. 2. ML classifiers' performance validation with MAE, MSE, and R^2.

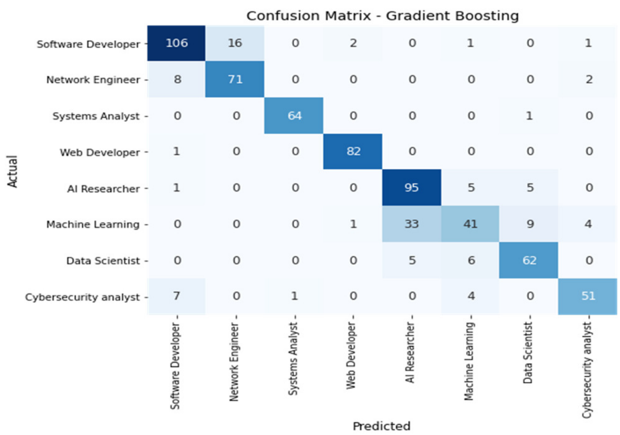

Fig. 3. Confusion matrix using GB.

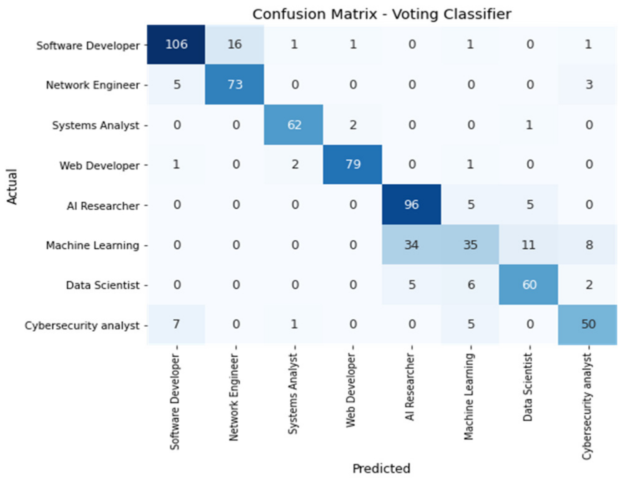

Fig. 4. Confusion matrix using the voting classifier.

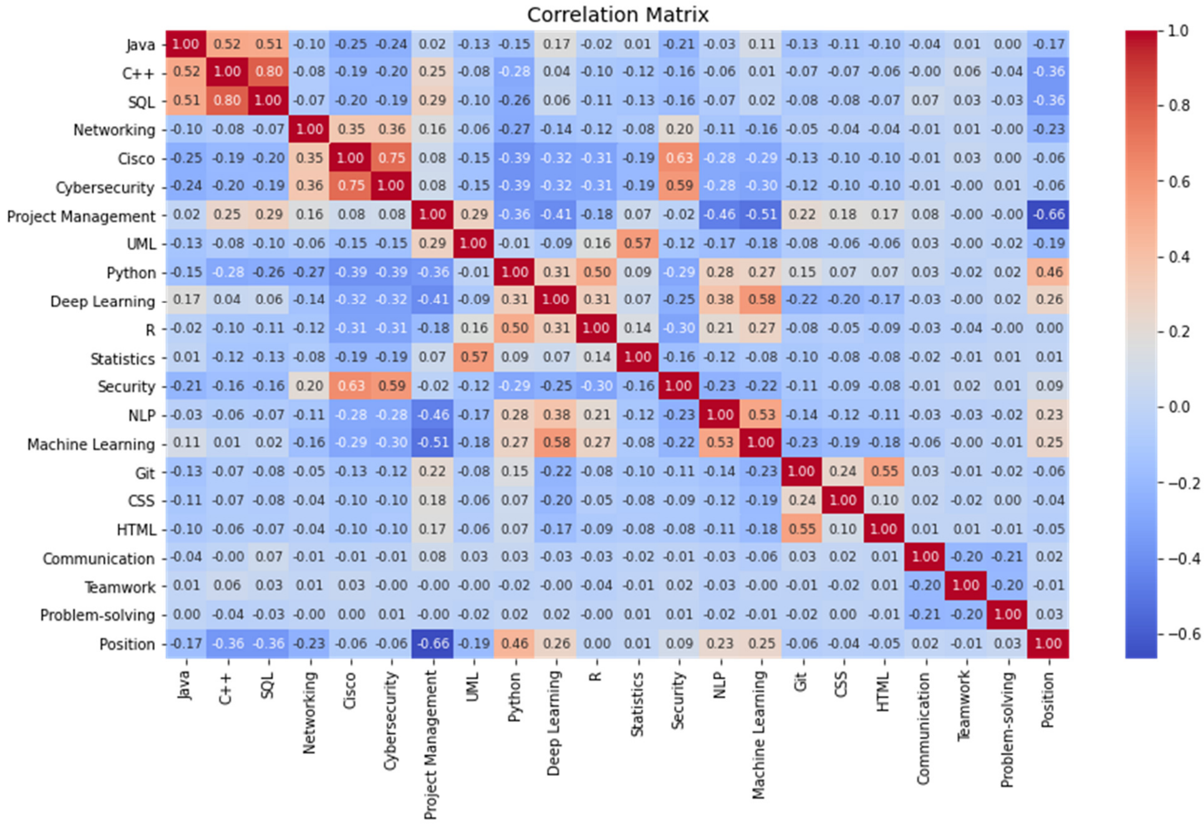

Fig. 5. Correlation matrix between the required skills and the job market.

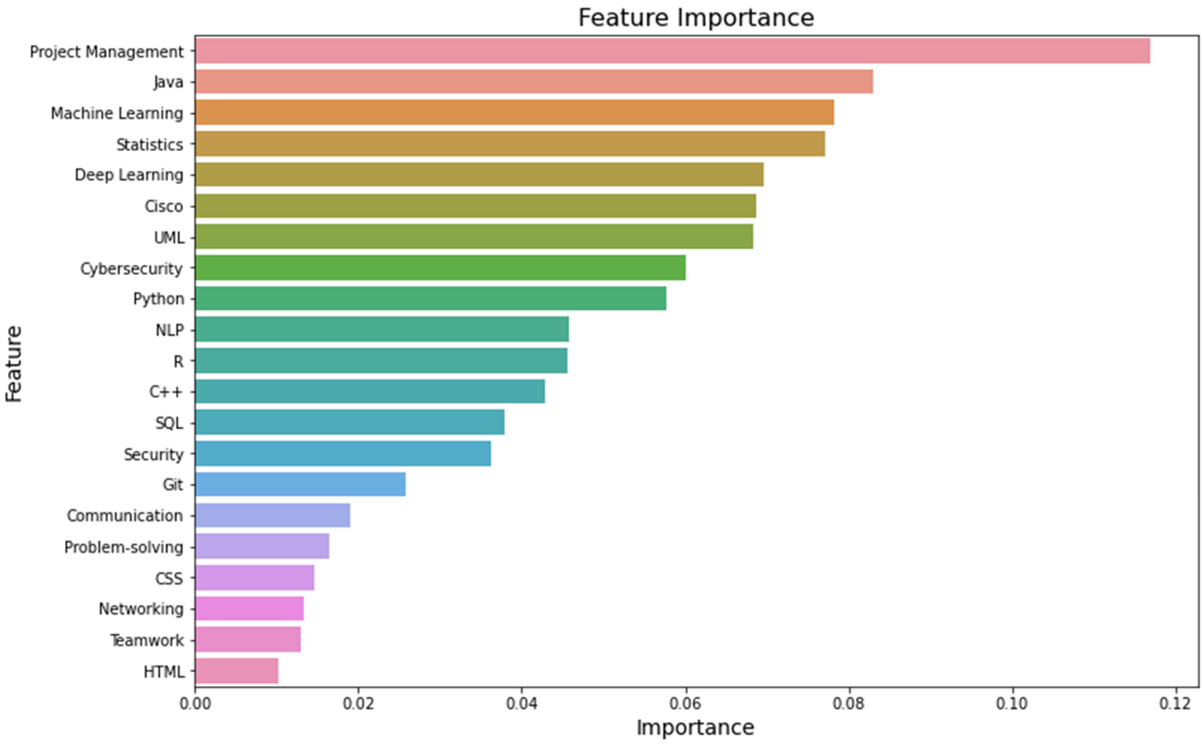

Fig. 6. Feature importance.

ANN was used in the second type of experiments. The considered model is an ANN designed for multi-class classification. It consists of an input layer, two hidden layers with ReLU activation functions, dropout layers to prevent overfitting and an output layer with a softmax activation function. The model is compiled employing the Adam optimizer with a learning rate of 0.0001 and the sparse categorical cross-entropy loss function. The experiment involves training the model on a dataset of computer science graduate students with various skills and positions. The dataset is divided into training and test sets, and the model is trained for 50 epochs with batch size of 32. The model's performance is evaluated on the test set, and accuracy is utilized as the main metric to assess the model's performance. Additionally, the training and validation accuracy and loss are plotted to visualize the training progress and identify any potential overfitting. The model performance in terms of accuracy recorded a score of 88%, as noticed in Figure 7. The loss function is presented in Figure 8.

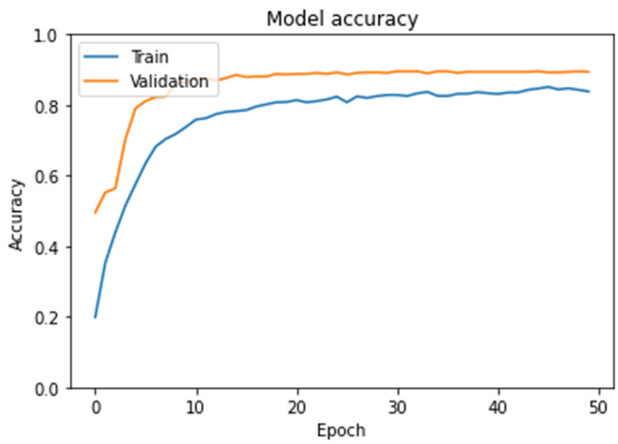

Fig. 7. Accuracy of the ANN.

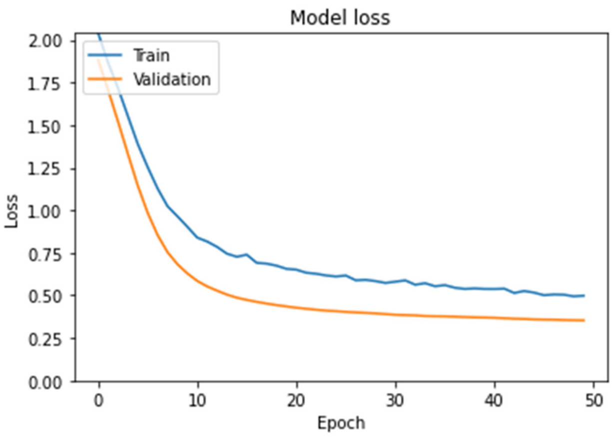

Fig. 8. Loss function of the ANN.

Overall, the DL model demonstrated strong performance, achieving accuracy, precision, and recall of 88%. This result indicates the DL model's effectiveness in the classification task compared with the other models. Overall, the computer science curriculum is required to be aligned with market needs through updating the course learning outcomes and organizing workshops or training courses for graduate students [11, 21, 28, 29]. Topics, such as ML, DL, Python programming, and cybersecurity, will help graduates close the skill gap with the market. The curriculum for computer science academic programs can incorporate these topics, and particular programming courses, or new courses can be designed.

VI. CONCLUSION

The current study addresses the issue of aligning graduate skills with the market demands, highlighting the discrepancy between the current curriculum offerings and the required job skills. By leveraging a novel dataset constructed from alumni surveys and job descriptions from hiring agencies, the current study trains ML and ANN models to classify computer science students' skills against market needs. The top five required skills identified—project management, Java, statistics, DL, and ML—underscore the need for curriculum revisions to better prepare graduates for the job market. This research sheds light on the challenges graduates face in finding suitable employment and emphasises the importance of adapting the educational programs to meet the industry needs. Additionally, the gap between graduate and market needs can be covered through updating the course learning outcomes by the required skills or through organising workshops and training courses.

ACKNOWLEDGMENT

The authors gratefully acknowledge the approval and the support of this research study by grant no. CSCR-2023-12-2285 from the Deanship of Scientific Research at Northern Border University, Arar, K.S.A.

REFERENCES

[1] M. Gawrycka, J. Kujawska, and M. T. Tomczak, "Competencies of graduates as future labour market participants–preliminary study," *Economic research - Ekonomska istrazivanja*, vol. 33, no. 1, pp. 1095–1107, 2020, https://doi.org/10.1080/1331677X.2019.1631200.

[2] W. M. S. Yafooz, E. A. Hezzam, and A.-H. M. Emara, "Machine learning based Collaborative Intelligent Closing Gap between Graduates and Labour Market Framework," in *International Conference on Artificial Intelligence and Smart Systems*, Coimbatore, India, Mar. 2021, pp. 1721–1726, https://doi.org/10.1109/ICAIS50930.2021.9395906.

[3] M. T. H. Duong, D. V. Nguyen, and P. T. Nguyen, "Using Fuzzy Approach to Model Skill Shortage in Vietnam's Labor Market in the Context of Industry 4.0," *Engineering, Technology & Applied Science Research*, vol. 10, no. 3, pp. 5864–5868, Jun. 2020, https://doi.org/10.48084/etasr.3596.

[4] H. Al-Dossari, F. A. Nughaymish, Z. Al-Qahtani, M. Alkahlifah, and A. Alqahtani, "A Machine Learning Approach to Career Path Choice for Information Technology Graduates," *Engineering, Technology & Applied Science Research*, vol. 10, no. 6, pp. 6589–6596, Dec. 2020, https://doi.org/10.48084/etasr.3821.

[5] V. Prikshat, A. Montague, J. Connell, and J. Burgess, "Australian graduates' work readiness – deficiencies, causes and potential solutions," *Higher Education, Skills and Work-Based Learning*, vol. 10, no. 2, pp. 369–386, Jan. 2020, https://doi.org/10.1108/HESWBL-02-2019-0025.

[6] S. Smith, E. Taylor-Smith, C. F. Smith, and G. Webster, "The impact of work placement on graduate employment in computing: Outcomes from a UK-based study," *International Journal of Work-Integrated Learning*, vol. 19, no. 4, pp. 359–369, 2018.

[7] S. Palmer, J. Coldwell-Neilson, and M. Campbell, "Occupational outcomes for Australian computing/information technology bachelor graduates and implications for the IT bachelor curriculum," *Computer Science Education*, vol. 28, no. 3, pp. 280–299, Jul. 2018, https://doi.org/10.1080/08993408.2018.1541385.

[8] I. Ruthotto, Q. Kreth, and J. Melkers, "Entering or advancing in the IT labor market: The role of an online graduate degree in computer science," *The Internet and Higher Education*, vol. 5, Oct. 2021, Art. no. 100820, https://doi.org/10.1016/j.iheduc.2021.100820.

[9] W. G. Alheadary, "Controlling Employability Issues of Computing Graduates through Machine Learning-Based Detection and Identification," *Engineering, Technology & Applied Science Research*, vol. 13, no. 3, pp. 10888–10894, Jun. 2023, https://doi.org/10.48084/etasr.5892.

[10] R. Boselli, M. Cesarini, F. Mercorio, and M. Mezzanzanica, "Classifying online Job Advertisements through Machine Learning," *Future Generation Computer Systems*, vol. 86, pp. 319–328, Sep. 2018, https://doi.org/10.1016/j.future.2018.03.035.

[11] D. Kumar, C. Verma, P. K. Singh, M. S. Raboaca, R.-A. Felseghi, and K. Z. Ghafoor, "Computational Statistics and Machine Learning Techniques for Effective Decision Making on Student's Employment for Real-Time," *Mathematics*, vol. 9, no. 11, Jan. 2021, Art. no. 1166, https://doi.org/10.3390/math9111166.

[12] A. Verma, K. Lamsal, and P. Verma, "An investigation of skill requirements in artificial intelligence and machine learning job advertisements," *Industry and Higher Education*, vol. 36, no. 1, pp. 63–73, Feb. 2022, https://doi.org/10.1177/0950422221990990.

[13] H. Huo *et al.*, "Predicting Dropout for Nontraditional Undergraduate Students: A Machine Learning Approach," *Journal of College Student Retention: Research, Theory & Practice*, vol. 24, no. 4, pp. 1054–1077, Feb. 2023, https://doi.org/10.1177/1521025120963821.

[14] E. Ramos-Monge, P. Fox, and A. Garcia-Piquer, "Addressing soft skill gaps in the digital employment market: the case of Spanish students in a technology-based university," *Education + Training*, vol. 65, no. 6/7, pp. 923–938, Jan. 2023, https://doi.org/10.1108/ET-04-2023-0165.

[15] B. Biasi and S. Ma, "The Education-Innovation Gap." National Bureau of Economic Research, Mar. 2022, https://doi.org/10.3386/w29853.

[16] P. Mohanamani and A. Latha, "Identifying Emerging Competencies demanded by the Employers from MBA Finance Graduates: Using NLP and TF - IDF Algorithm," in *7th International Conference On Computing, Communication, Control And Automation*, Pune, India, Aug. 2023, pp. 1–6, https://doi.org/10.1109/ICCUBEA58933.2023.10392227.

[17] M. Rakhimov, A. Yuldashev, and D. Solidjonov, "The role of artificial intelligence in the management of e-learning platforms and monitoring knowledge of students," *Oriental renaissance: Innovative, educational, natural and social sciences*, vol. 1, no. 9, pp. 308–314, 2021.

[18] M. Aldayel, M. Ykhlef, and A. Al-Nafjan, "Deep Learning for EEG-Based Preference Classification in Neuromarketing," *Applied Sciences*, vol. 10, no. 4, Jan. 2020, Art. no. 1525, https://doi.org/10.3390/app10041525.

[19] L. Benhayoun and D. Lang, "Does higher education properly prepare graduates for the growing artificial intelligence market? Gaps' identification using text mining," *Human Systems Management*, vol. 40, no. 5, pp. 639–651, Jan. 2021, https://doi.org/10.3233/HSM-211179.

[20] F. K. Abbasi, A. Ali, and N. Bibi, "Analysis of skill gap for business graduates: managerial perspective from banking industry," *Education + Training*, vol. 60, no. 4, pp. 354–367, Jan. 2018, https://doi.org/10.1108/ET-08-2017-0120.

[21] G. Li, C. Yuan, S. Kamarthi, M. Moghaddam, and X. Jin, "Data science skills and domain knowledge requirements in the manufacturing industry: A gap analysis," *Journal of Manufacturing Systems*, vol. 60, pp. 692–706, Jul. 2021, https://doi.org/10.1016/j.jmsy.2021.07.007.

[22] M. U. Hassan, S. Alaliyat, R. Sarwar, R. Nawaz, and I. A. Hameed, "Leveraging deep learning and big data to enhance computing curriculum for industry-relevant skills: A Norwegian case study," *Heliyon*, vol. 9, no. 4, Apr. 2023, Art. no. e15407, https://doi.org/10.1016/j.heliyon.2023.e15407.

[23] E. Karakolis *et al.*, "Bridging the Gap between Technological Education and Job Market Requirements through Data Analytics and Decision Support Services," *Applied Sciences*, vol. 12, no. 14, Jan. 2022, Art. no. 7139, https://doi.org/10.3390/app12147139.

[24] B. Mahesh, "Machine Learning Algorithms - A Review," *International Journal of Science and Research*, vol. 9, no. 1, pp. 381–386, Jan. 2019, https://doi.org/10.21275/ART20203995.

[25] W. M. S. Yafooz, A.-H. M. Emara, and M. Lahby, "Detecting Fake News on COVID-19 Vaccine from YouTube Videos Using Advanced Machine Learning Approaches," in *Combating Fake News with Computational Intelligence Techniques*, M. Lahby, A.-S. K. Pathan, Y. Maleh, and W. M. S. Yafooz, Eds. New York, NY, USA: Springer, 2022, pp. 421–435.

[26] D. Burka, C. Puppe, L. Szepesvary, and A. Tasnadi, "Voting: A machine learning approach," *European Journal of Operational Research*, vol. 299, no. 3, pp. 1003–1017, Jun. 2022, https://doi.org/10.1016/j.ejor.2021.10.005.

[27] R. Yacouby and D. Axman, "Probabilistic Extension of Precision, Recall, and F1 Score for More Thorough Evaluation of Classification Models," in *First Workshop on Evaluation and Comparison of NLP Systems*, Nov. 2020, pp. 79–91, https://doi.org/10.18653/v1/2020.eval4nlp-1.9.

[28] W. M. S. Yafooz, E. A. Hizam, and W. A. Alromema, "Arabic Sentiment Analysis on Chewing Khat Leaves using Machine Learning and Ensemble Methods," *Engineering, Technology & Applied Science Research*, vol. 11, no. 2, pp. 6845–6848, Apr. 2021, https://doi.org/10.48084/etasr.4026.

[29] C. Teichert, I. Liefner, and A. Otto, "How wide is the gap? Comparing geography graduates' labor market success with that of peers from business and computer science," *Journal of Geography in Higher Education*, vol. 46, no. 4, pp. 599–627, Oct. 2022, https://doi.org/10.1080/03098265.2021.1960490.

[30] N. E. A. M. Almi, N. A. Rahman, D. Purusothaman, and S. Sulaiman, "Software engineering education: The gap between industry's requirements and graduates' readiness," in *IEEE Symposium on Computers & Informatics*, Kuala Lumpur, Malaysia, Mar. 2011, pp. 542–547, https://doi.org/10.1109/ISCI.2011.5958974.