UDC 004.032.26
DOI: 10.15587/1729-4061.2026.357493

IMPROVING THE EFFICIENCY AND ACCURACY OF CLASSIFICATION OF RECIPIENTS OF THE FREE NUTRITION FOOD PROGRAM THROUGH THE APPLICATION OF KNOWLEDGE DISTILLATION IN A CONVOLUTIONAL NEURAL NETWORK ALGORITHM

Relita Buaton
Corresponding author
Doctor of Computer
Department of Information Technology*
E-mail: relitabuaton@kaputama.ac.id
ORCID: https://orcid.org/0000-0003-0968-0017
Mesra Betty Yel
Doctor of Computer
Department of Information System
STIKOM Cipta Karya Informatika
Radin Inten II str., 8 5, RT.5/RW.14, Duren Sawit, Kec. Duren Sawit, Kota Jakarta Timur, Daerah Khusus Ibukota Jakarta, Indonesia, 13440
ORCID: https://orcid.org/0009-0002-8336-5058
Novriyenni
Master of Computer
Department of Information System*
ORCID: https://orcid.org/0000-0001-7511-0149
Anton Sihombing
Master of Management
Department of Management Information System*
ORCID: https://orcid.org/0000-0001-8652-198X
Ida Ria Royentina Sidabukke
Master of Health
Department of Midwifery
Universitas Sari Mutiara Indonesia
Kapten Muslim str., 79, Helvetia Tengah, Kec. Medan Helvetia, Kota Medan, Sumatera Utara, Indonesia, 20124
ORCID: https://orcid.org/0009-0007-6544-3178
*STMIK Kaputama
Veteran str., 4A-9A, Tangsi, Kec. Binjai Kota, Kota Binjai, Sumatera Utara, Indonesia, 20714

The object of this study is a deep learning-based classification system applied to recipients of the Free Nutritional Food Program, using recipient data as a representation of the eligibility determination process in providing social assistance. The problem to be solved is the high computational complexity of large-capacity convolutional neural networks (CNN) models which, despite their high accuracy, require significant computational resources and are therefore less than optimal for large-scale implementation. To overcome this, this study applies the Knowledge Distillation method, utilizing a large-capacity CNN as the teacher model and a lightweight architecture CNN as the student model through soft label-based knowledge transfer. According to this study, it is shown that the student model generated by distillation is about 90–93% accurate. Also, this figure is very close to that of the teacher model (from 92–95%) and much better than that of a CNN without distillation model (85–88%). This is an improvement since this distillation method can transfer information in the form of richer probabilities than simply hard labels as done do in traditional training. The model proposed in this work has many advantages such as higher accuracy, more compact size, and faster inference times. These features help in making the classification process computationally less intensive. Furthermore, this leads to more efficient memory use and lower energy consumption. These results could be applied in many deep learning classification systems, particularly resource limits devices. Such application is observable in the implementation of Free Nutritious Food Program under real life conditions which requires better accuracy with no loss on efficiency

Keywords: Food Program, CNN, student distillation model, classification system

Received 04.02.2026
Received in revised form 30.03.2026
Accepted 13.04.2026
Published 30.04.2026

***How to Cite:** Buaton, R., Yel, M. B., Novriyenni, N., Sihombing, A., Sidabukke, I. R. R. (2026). Improving the efficiency and accuracy of classification of recipients of the free nutrition food program through the application of knowledge distillation in a convolutional neural network algorithm. Eastern-European Journal of Enterprise Technologies, 2 (3 (140)), 37–45. https://doi.org/10.15587/1729-4061.2026.357493*

1. Introduction

This program, named Free Nutritious Food Program is a policy from the Indonesian government to improve human resources quality and this program has been applied to make sure school-age children receive adequate nutrients. It is expected that the healthier and more productive generation with academic skills will be trained through this program [1, 2].

Nonetheless, although some factors have already been taken into account in its application (e.g. rejection by recipients), there are still several challenges, including incompatibility with local dietary habits in certain regions, rejection on the part of recipients and lack of public knowledge about healthy eating habits [3, 4]. The problem studied reported in this paper is related to the distribution of assistance of the Free Nutritious Food Program, which needs an exact and timely classification for correct identification of beneficiaries. Actually, there are still some challenges for beneficiary selection process, including low accuracy when using heterogeneous data and inference performance bottleneck when querying complex artificial intelligence model. A popular technique is convolutional neural network (CNN), which has been widely used in image data processing. CNN: It is an artificial neural network that performs feature extraction and recognition through convolutional layers which yields good performance in classification tasks. Nevertheless, this approach typically demands a lot of computing power, substantial processing time, and high-end hardware. A difficulty manifests when many such systems are deployed in distributed facilities with limited technology and the need for quick processing of data [5, 6].

To overcome the aforementioned concerns, this work proposes the knowledge distillation method to be incorporated with CNNs which will lead to a more computationally-efficient process while maintaining accuracy [7, 8]. This approach utilizes knowledge distillation, where a student model learns to mimic the output probabilities of a larger and high-capacity model known as teacher model. In this approach, it assumes that the student model has similar accuracy with a smaller model size, faster processing speed and fewer resources than teacher [9, 10]. This is done in practice by using two models of gradually decreasing complexity, the teacher model and the student model [11, 12]. First, to achieve high accuracy in identifying program recipients, the teacher model is trained. Then, this method is using soft labels [13, 14] to transfer knowledge from the teacher model into the student model. Loss from the original labels and loss from distillation results are combined to then train student model. This technique enables the model to preserve an accuracy yet consumes less computational resources and inference time [15].

Therefore, research on developing classification models to improve efficiency in the free nutritious food program through the application of knowledge distillation in the convolutional neural network (CNN) algorithm is crucial. This approach enables more accurate and targeted identification and grouping of beneficiaries based on their actual characteristics and conditions. Knowledge distillation has the potential to produce more sophisticated models with lower complexity, faster inference times, and more efficient computational resource requirements, without significant performance degradation. Thus, the resulting models not only maintain adequate classification rates but are also more adaptable for implementation on large-scale systems and devices with limited resources, thus supporting effective and sustainable program distribution optimization.

Therefore, research focused on developing a knowledge distillation model using a CNN algorithm is relevant in the context of the classification problem of free nutritional food program recipients. This is because the combination of knowledge distillation with a CNN algorithm can improve classification efficiency and accuracy by learning from the knowledge distillation parameters.

2. Literature review and problem statement

Study [16] discusses feature maps to be used as knowledge distillation from layers that will form vectors into one dimension, then the information extraction process is carried out so that the training process can read data effectively. The results of the application are in the form of accuracy with a value of 90% which is able to read data optimally, there are several shortcomings in the model when integrating the layers to maximize the model. So the application of knowledge distillation first before the training model can help the integration.

Study [17] presents the results of the application of knowledge distillation combined with a simpler process between the teacher architecture and the student architecture in modeling and reading data patterns from small amounts and large amounts of data that greatly affect the accuracy of the model in making predictions. This approach creates a gap in terms of optimal hyperparameter assessments so that the classification is not the same as the actual data. This problem can be overcome by implementing Bayesian optimization on the CNN architecture which is then integrated with knowledge distillation so that the architecture can learn data patterns and increase the accuracy of the teacher model. So the integration of knowledge distillation can be a solution in the CNN architecture.

Study [18] presents a research on the influence of dataset characteristics on the effectiveness of knowledge distillation in classifying images, showing that dataset size, data distribution complexity and domain bias affect the successful transfer of knowledge from the teacher model to the student model with an accuracy of 89% however, there are some unresolved issues related to the lack of standard guidelines for selecting the optimal dataset for the process of transferring information to the distillation architecture to be processed in the CNN algorithm. To overcome this difficulty, a standard evaluation framework for datasets in distillation scenarios was developed, but variations in data characteristics remain difficult to control. All this suggests that it is advisable to conduct studies on the application of knowledge distillation in CNN algorithms with dataset characteristics being considered as a key factor in the experimental design.

Study [19] presents a research on the development of a lightweight and easily interpretable deep learning model for nutritional analysis in mobile-based health applications by utilizing knowledge distillation, depthwise separable convolution, and attention mechanisms. The results of this study indicate that the distillation process can reduce model size by more than 60% while increasing energy efficiency without significantly reducing accuracy. This condition is mainly influenced by the size and complexity of CNN models, which are generally large, require high computing power, and require significant memory for both training and operation. On devices with limited specifications such as mobile devices or edge devices, optimal implementation of CNNs is a significant challenge. One approach to overcome this problem is to incorporate lighter and more adaptive AI methods, although the computational requirements for generating model explanations remain a challenge. Overall, these findings indicate that further studies are needed on the application of knowledge distillation in CNN algorithms to support the implementation of more efficient and explainable models on mobile devices.

Study [20] explains that knowledge distillation integrated with CNN algorithms can improve CNN interpretability with the feature extraction process into the model architecture to read patterns based on patterns from teacher models and student models. Accuracy in this study is 87% in detecting diseases based on medical data, in the process of reading patterns there are unresolved problems related to model accuracy that must be adjusted to the knowledge distillation parameters in dealing with big data which generally must have a lot of memory so that models that can train and make predictions on big data must be integrated with knowledge distillation so that they are computationally efficient.

Study [21] presents a research on the implementation of CNN and RNN in detecting food safety and predicting Risk based on image and time series data. It shows that CNN excels in visual feature extraction to detect defects or contamination, while RNN/LSTM excels in time series analysis to predict food safety risks. However, there are unresolved issues related to large data requirements, high computational system complexity and low model interpretability in the context of CNN architecture distillation. Knowledge distillation, as one, as one of the approaches to reduce the size and complexity of CNN models without sacrificing performance, one way to overcome these difficulties is to integrate lightweight models and explainable AI approaches, but the computational challenges remain significant for real-time systems. All these indicate that it is advisable to conduct studies on the implementation of knowledge distillation in CNN algorithms to produce more efficient food safety detection models that can be implemented in IoT-based systems.

Study [22] will explain that in the application of MobileNetV3 for food identification on smartphone devices, it is processed and combined with an ensemble method so that it will increase the accuracy in carrying out detection which will be processed in training with the addition of Grad-CAM++ feature extraction so that it produces 90% accuracy in food recognition detection, but in the process there is a sharpness that requires large computation to process large amounts of data. The solution in this case must be integrated with knowledge distillation so that computation can be reduced to save memory.

3. The aim and objectives of the study

The aim of this study is to implement the knowledge distillation method on a CNN algorithm for accuracy in classifying beneficiaries of nutritious food programs.

In pursuit of this aim the following actions were taken:

– cross-relationship of CNN algorithm architecture with knowledge distillation;

– assessing the model in the course of classification.

4. Materials and methods

In this section the researcher focuses this study on building a deep learning-based classifier to check whether Free Nutritious Food Program beneficiaries are eligible or not. It is advised that the candidate recipient data will be mainly focused in this decision-making system. But the subject of the study itself might be deemed less relevant, as the main aim is restated previously. Here, the researcher assumes that given the recipient data used has been preprocessed and filtered initially, it is representative of the actual falls in field conditions for practical evaluation of models. It is also assumed that the current structured labeling itself is valid and accurate enough to represent the eligibility of each individual. In addition, all the dataset is divided into training, validation and testing in a balanced way to ensure that model evaluation can be performed properly. It is also assumed that the Knowledge Distillation method can be used for this particular kind of social assistance classification.

The primary hypothesis being put forth is that implementation of the knowledge distillation method on a CNN model can help enhance classification performance along with decreasing computational complexity. More specifically, the researcher intends to generate a model that is more efficient than an ordinary CNN without greatly forgoing accuracy. The researchers perform a relevant parameters-driven classification task and are limited by major metrics: F1-score, precision, and accuracy that the authors suggest can accurately measure model performance. It also seeks to establish whether Free Nutritious Food Program beneficiaries qualify for assistance. As such, prospective recipients' data is the main input to the system when deciding. This branch of the study may be viewed as rather basic, as it already points to the most significant influences discussed earlier in this paper, but is still vital for systematic assessment.

This study utilized the data provided by the recipients (after basic preprocessing) to simulate real-world situations and to check whether or not this model is accurate. Assuming that the structured labels already exist, these are believed to be trustworthy and useful in terms of them reflecting whether or not an applicant is eligible on their own. The data is also split into training, validation and test sets evenly to ensure the model evaluation is performed appropriately. Also, it is assumed that the knowledge distillation method can be successfully implemented for this type of classification of social assistance. The objective of this study is to initially introduce the hypothesis that, on one hand, knowledge distillation applies and increases the performance rate of CNN classification model while decreasing computational complexly. To put it simply, the researchers say they want to create a smaller architecture than an ordinary CNN but achieve nearly the same accuracy. The Matplotlib library is used to visualize results.

All data processing and model development were conducted in a Linux-based development environment, with environment management software such as Jupyter Notebook for ease of experimentation and documentation. On the hardware side, this study utilized servers with multi-core processors and GPUs (Graphics Processing Units) that support computational acceleration to expedite the model training process. The GPUs used had sufficient memory capacity to handle large CNN models, which was essential for running experiments with large datasets. Furthermore, data was stored on high-speed hard drives to ensure fast data access during the model training and evaluation process. The tolerance of this study encompasses several aspects that must be considered in model implementation and evaluation. Tolerance relates to model accuracy, where the model is expected to achieve a predetermined minimum level of accuracy, with an acceptable margin of error based on evaluation standards such as the F1-score and recall. The architectural design of this study is presented in the form of a flow diagram to illustrate the overall stages of knowledge distillation implementation in CNN, as shown in Fig. 1.

Fig. 1 shows the research flow for applying knowledge distillation to the convolutional neural network (CNN) algorithm to improve the efficiency and accuracy of Free Nutritional Food Program recipient classification. The research process begins with data collection from the National Nutrition Agency, which provides data related to potential program recipients. This data serves as the primary basis for developing the classification system. The next stage is data processing, which includes three main processes: data cleaning, data transformation, and feature engineering. In the data cleaning stage, incomplete data, duplication, and anomalies are removed to maintain data quality. Next, data transformation is performed to adjust the data format and scale so that it can be processed by the CNN model. The feature engineering stage aims to extract and select relevant features to make the information used in the classification process more representative.

Once the data is ready, the process proceeds to the knowledge distillation stage, which involves two main models: the teacher model and the student model. The teacher model is a large-capacity CNN model trained to achieve optimal classification performance. Knowledge from the teacher model is then transferred to the student model through soft labels using a distillation loss, while hard labels from the original data are retained to calculate the classification loss. The combination of these two types of losses allows the student model to learn more informative prediction patterns without requiring high complexity. Now, the next step is to implement the CNN architecture, which consists of several parameters such as input size, number of filters, pooling layers, activation functions and fully connected layers. These parts are involved in feature extraction and the classification based on that distilled data. While this model should be more compact and much faster to infer, it will still maintain a high accuracy. The last phase of this study is the evaluation that measures how well the classification model works. The assessment is based on a few measures including accuracy, precision, recall and F1-score are used to evaluate how well the model can classify beneficiaries of Free Nutritious Food Program correctly and efficiently. All in all, the study workflow is depicted as a flow chart that delivers the systematic tasks/process from data selection and collection to modeling and evaluation.

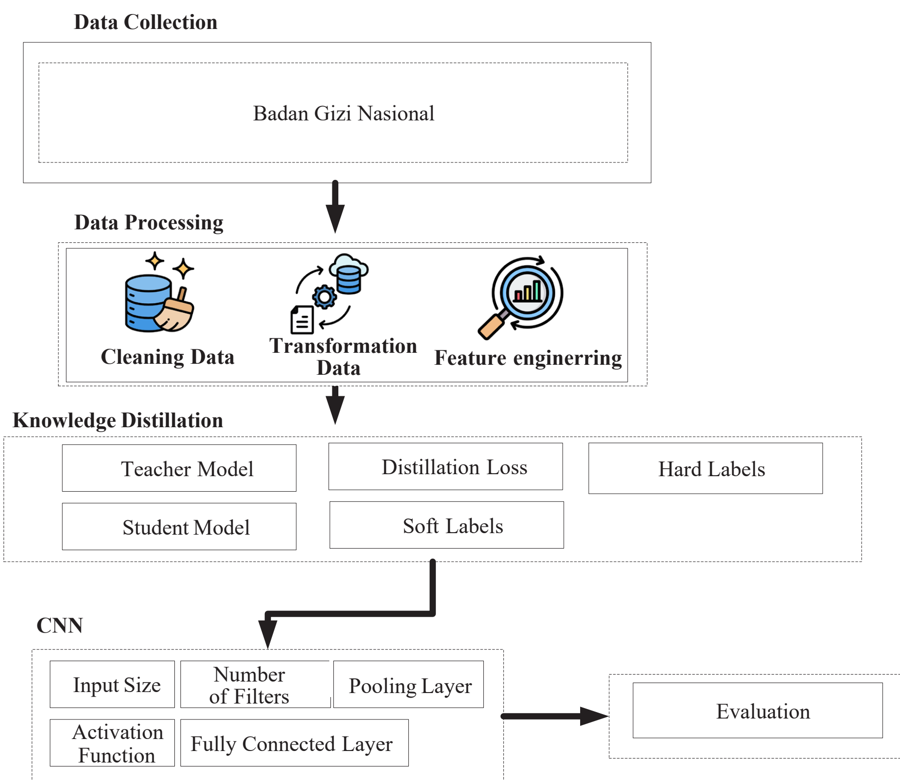

Fig. 1. System architecture

5. Results of applying knowledge distillation to CNNs in the Free Nutritious Food Program

5. 1. Implementation of knowledge distillation integration in CNN architectures

The theoretical framework of this study illustrates how the Free Nutritious Food Program would affect students' academic achievement. Directly, as an increased nutritional status of the students; and indirectly, as higher attendance and motivation to learn. Other factors such as SES and students' prior achievement are proposed to moderate this relationship. The Free Nutritious Food Program as independent variable and it is estimated to have positive effect in students' learning outcomes. Nutrition works as a mediatory variable so that the availability of nutritional intake can improve students' body health, concentration and cognitive function leading to improving their academic performance. The dependent variable has been academic achievement, proxied by report card grades of students and so is the independent variable that attention to their nutrition condition. On the other hand, socio-economic status and previously achieved academic performance can make intervention effects stronger or weaker.

The relationships between the variables can be depicted in a diagram such as: Free Nutritious Food Program → Nutritional status → Academic Achievement; SES and prior academic performance are moderating variables.

In this study, multiple machine learning methods were integrated during the model optimization process to study the relationship between nutrient intake and improved academic performance. The relationship patterns between nutritional components covered (carbohydrates, fiber, protein, fat and energy) that can bring better academic performance were identified using the Association Rule Mining method with Apriori algorithm. And using a K-Means clustering algorithm to cluster students based on their performance patterns before and after the program, as well as nutritional profiles. On the other hand, a implemented a classification model based on CNN to predict new students' academic performance based on previously trained data. The result of the analysis can be easily understood, for example if to find a rule that states eating high protein increases your English grade then it is possible to modify the menu by adding food with high protein. The clustering results also help identify the groups of students who benefit most from the program, enabling more targeted interventions. Moreover, it allows to predict future students' academic records based on their nutrition profiles. This research analysis aimed to identify relationships between nutrition and enhanced performance, cluster students based on their program responses, and predict academic outcomes for the subsequent period with respect to those clusters. These outputs are in the form of association rules with their support, confidence and lift measures, student profiles for each cluster as well as nutritional recommendations which could enhance academic performance in different subjects.

Considering the CNN algorithm, by using knowledge distillation approach to formulate the model, its performance with respect to metrics e.g., accuracy precision recall and computational effectiveness among Free Nutritious Food Program beneficiaries is shown in Table 1.

Table 1

Association rules for student grades before and after the Free Nutritious Food Program

No.	Antecedence	Consequence	Support	Confidence
1	Social (Odd Semester) = B+	Social (Even Semester) = A+	0.396767	0.8
2	English (Odd Semester) = B+	English (Even Semester) = A+	0.332109	0.8
3	Mathematics (Even Semester) = B+	Religion (Even Semester) = A	0.337987	0.8
4	Mathematics (Even Semester) = B+	Social (Even Semester) = A	0.400441	0.8
5	Religion (Even Semester) = A	Social (Even Semester) = A+	0.332109	0.8
6	Social (Even Semester) = A	Social (Even Semester) = A+	0.332109	0.8

Best-rule summary:

– Social studies improves from B+ → A+ with support ≈ 39% and confidence 80%;

– English improves from B+ → A+ with support ≈ 33% and confidence 80%;

– Religion = A is associated with social studies = A+ with support ≈ 33% and confidence 80%;

– Mathematics = B+ is associated with higher outcomes in English/social studies (A or A+) with support ≈ 33% and confidence 80%;

– No significant association indicates a direct improvement in mathematics itself.

The results of applying knowledge distillation to the convolutional neural network (CNN) algorithm indicate that the student model is able to maintain a classification accuracy level close to that of the teacher model, while simultaneously achieving higher computational efficiency in the process of identifying beneficiaries of the Free Nutritional Food Program. The distillation process through the transfer of soft targets and temperature scaling has been proven to improve the model's generalization ability on test data. The results also show that the program is positively associated with post-program improvements in subjects emphasizing verbal, social, and attitudinal competencies, namely social studies, English, and religious education. In contrast, no strong evidence supports direct improvement in Mathematics, suggesting that numerical skills and formal reasoning may require additional structured pedagogical interventions to achieve comparable gains. Overall, the support and confidence profile demonstrates that the Free Nutritional Food Program correlates with meaningful academic improvements in non-STEM areas, while Mathematics may benefit from complementary instructional strategies. The following section provides a description of each cluster:

1. Cluster of average scores by education level: pre- vs post-program.

To measure the program's impact at each educational level, student scores before and after the implementation of a Free Nutritious Food Program were analyzed using K-Means for grouping students in terms of similarities in their achievement patterns. As shown in Table 2, the clustering results show an effective positive effect on each of its group, but with different degree of improvement.

Table 2

Centroids of average student scores (pre/post) by education level

Cluster	Education level (centroid)	Pre-program avg	Post-program avg
1	2.220713	90.87521	92.51072
2	3.693837	82.91327	84.97662
3	1.527881	82.26082	84.47545

Cluster 1 (≈ junior high school): the average rose from 90.87 to 92.51 – signifying that it had the greatest impact in response to the program.

Cluster 2 (≈ senior high school & vocational high school): increase from 82.91 to 84.98 – indicating moderate rise.

Cluster 3 (≈ elementary school): increased from 82.26 to 84.48 – the smallest increase of all clusters. All of the clusters experienced increases post-intervention, but the most pronounced upward trend was located at the junior high school level (SMP) (Fig. 2).

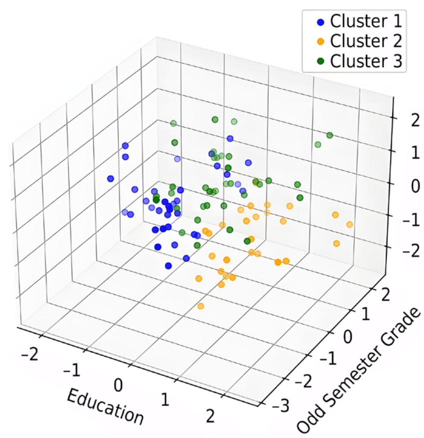

Fig. 2. Clustered averages (pre vs post) by education level

Based on code interpretation, it is possible to show that the distinguish aspect for Cluster 1 (high achievers) witnessed a steady incremental improvement. Cluster 2 was relatively successful and included many students. On the other hand, Cluster 3 started with lower initial scores and showed gradual improvement only. This means that the effect is mostly positive but the size of this effect greatly depends on its students' baseline level and initial ability (Table 3).

Table 3

Centroids for mathematics (pre/post) and education level

Cluster	Math (odd)	Math (even)	Education level
1	9.795414462	9.913580247	1.185185185
2	8.76673428	8.959432049	2.795131846
3	8.621262458	8.940199336	0.551495017

Cluster 1 (top-performing, ≈ junior high school level).

Has the top math scores pre (9.80) and post (9.91). There is a distinct improvement, and the data distribution is rather tight. This means that junior high school students had the highest responsiveness to the program, which is consistent with early adolescence being an optimal cognitive development stage when supported by adequate nutrition.

Cluster 2 (medium level, ≈ secondary/technical education).

Ratings increased from 8.77 to 8.96. But a broader spread of data shows other factors beyond nutrition, like study habits and curriculum workload, also influence learning outcomes.

Cluster 3 (grade-level, ≈ elementary school).

Has the least average score (8.62 → 8.94) but with high variability in data. This means that, while at this level these students still need nutritional support, they also require a more suitable learning modality.

Overall, all clusters show improvement. However, there is often less improvement observed in mathematics than language and social studies. Its impact of the program is most apparent at junior high school level fairly apparent at senior high school/vocational school level, and least noticeable in elementary school. This is followed by a clustering graph of mathematics (odd and even semesters, educational levels in Fig. 3).

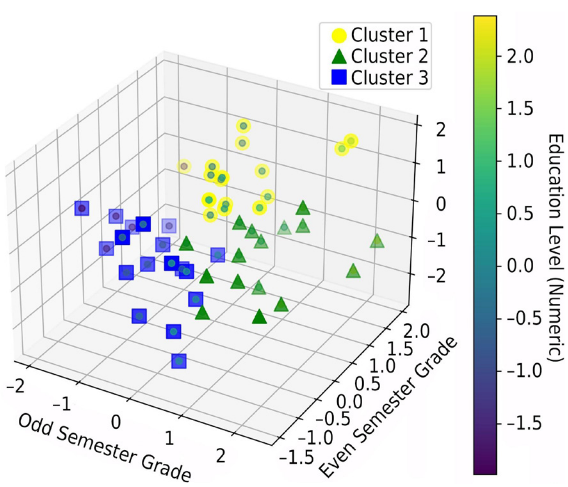

Fig. 3. 3D clustering for mathematics (odd, even, level)

Cluster 1 (green, junior high school level) shows higher and stable performance with the most gain. Cluster 2 (blue-green, comparable to high school/vocational school) is intermediate level, with quite steady improvement, and this is where the curriculum and strategies for learning also matters. On the other hand, Cluster 3 (yellow, comparable to elementary school) is already at a low baseline and has limited improvement because results are still largely determined by basic skills such as early numeracy and pedagogical approaches. The resulting clusters formed from multiple subjects (religion, mathematics, English, and social studies) organized into a shift on the odd semester to even semester.

Table 4

Centroids (pre-program, odd semester)

Cluster	Religion	Mathematic	English	Social
0	9.02020202	9.018518519	9.001683502	9.052188552
1	9.746031746	9.819444444	9.841269841	9.609126984
2	8.806896552	7.675862069	7.731034483	8.448275862
3	8.264705882	8.882352941	8.460784314	7.274509804
4	5.875	9.125	9.4375	9.5

Table 5

Centroids (post-program, even semester)

Cluster	Religion	Math	English	Social
0	10	9.884615385	10	9.947963801
1	9.442073171	9.737804878	9.237804878	9.743902439
2	8.924564797	8.845261122	8.945841393	8.893617021
3	4	10	9.6	10
4	8.109375	7.703125	7.96875	7.8125

Comparative analysis.

The top group improved ever more, where Pre-Cluster 1, previously attaining a near-perfect score for every subject (including both Religious knowledge and English) achieved Post-Cluster 0 with scores of 10 in Religion and English and very high marks in Maths & Social Studies. And that represents an improvement among those highest-performing group. This study applied a convolutional neural network (CNN) algorithm specifically the knowledge distillation method to classify Free Nutritious Food Program recipients based on socioeconomic conditions and other supportive factors. The teacher model shares its knowledge with student models so they can run more efficiently without sacrificing accuracy. After the aforementioned configurations, model performance was assessed by a confusion matrix and inference time in order to assess computational efficiency gains (refer to Fig. 4).

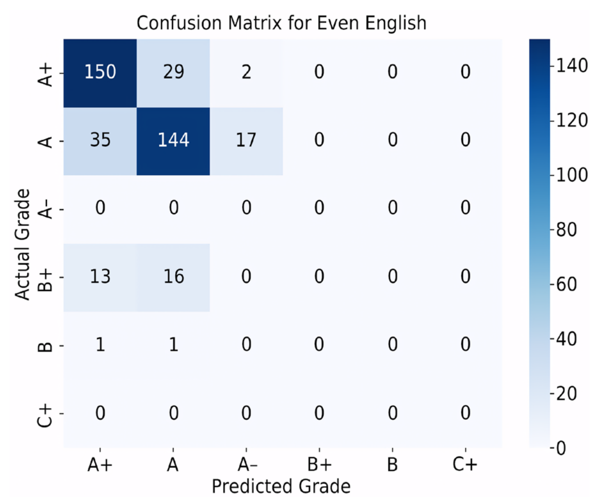

Fig. 4. Confusion matrix for English (even semester)

An interpretation of Fig. 4 shows that the overall accuracy rate is in the range of 80–85%. The highest precision and recall values are found in categories A+ and A. Classification errors occur most frequently between categories A and A+, due to the very close cutoff values at the highest level. These results are also consistent with the findings from the A-Priori method, which indicate an improvement in the high-value categories as a result of the implemented program.

5. 2. Solving the model for classification evaluation

The mathematical formulation utilized here was to evaluate the accuracy of the model in distinguishing between recipients and nonrecipients of Free Nutritious Food Program. The evaluation was conducted to check how much the model generated by applying knowledge distillation can retain or even increase its classification accuracy while improving on computational efficiency in relation to the teacher model.

The classification of program recipients is mathematically framed as a supervised classification problem, where the goal is to systematically and quantitatively map the characteristic data of potential recipients into eligibility categories. This problem can be expressed as a dataset as in the following equation

$$\mathcal{D}=\left\{\left(x_i, y_i\right)\right\}_{i=1}^{N}, \tag{1}$$

where $x_i \in \mathbb{R}^d$ represents the receiver feature vector and $y_i \in \{1, 2, \ldots, K\}$ is the feasibility label. The learning objective is to find the nonlinear mapping function which minimizes empirical risk and is able to classify data accurately. The Convolutional Neural Network (CNN)-based teacher model is defined as

$$z^T = f_T\left(x;\theta_T\right), \tag{2}$$

with Z^T Logit output before the softmax activation function. The prediction probability is calculated using the softmax function in equation (3)

$$P_T^{(k)} = \frac{\exp\left(z_k^T\right)}{\sum_{j=1}^{K}\exp\left(z_j^T\right)}. \tag{3}$$

Then the teacher model is trained by minimizing the cross-entropy loss which can increase the classification accuracy in equation (4)

$$\mathcal{L}_T = \sum_{k=1}^{K} y_k \log p_T^{(k)}. \tag{4}$$

Next, the lighter student model will be processed in equation (5)

$$z^s = fs\left(x;\theta s\right). \tag{5}$$

In the knowledge distillation approach, the probability distribution is softened using the temperature parameter $T > 1$ contained in equation (6), (7):

$$p_T^{(k)}\left(T\right)=\frac{\exp\left(z_K^T / T\right)}{\sum_{j=1}^{k}\exp\left(z_j^T / T\right)}, \tag{6}$$

$$p_S^{(k)}\left(T\right)=\frac{\exp\left(z_K^S / T\right)}{\sum_{j=1}^{k}\exp\left(z_j^S / T\right)}. \tag{7}$$

Distillation loss is defined using Kullback-Leibler divergence

$$\mathcal{L}_{KD} = T^2\sum_{k=1}^{K} p_T^{(k)}\left(T\right)\log\frac{p_T^{(k)}\left(T\right)}{p_S^{(k)}\left(T\right)}. \tag{8}$$

The total objective function optimized by the student model is a combination of classification loss and distillation loss:

$$\mathcal{L}_{total}=\alpha\mathcal{L}_{CE}+\left(1-\alpha\right)\mathcal{L}_{KD}, \tag{9}$$

$$\mathcal{L}_{CE} = -\sum_{k=1}^{K} yk \log p_S^{(k)}. \tag{10}$$

The student model parameters are updated using a gradient-based optimization algorithm

$$\theta_S^{(t+1)} = \theta_S^{(t)} - \eta\nabla_{\theta s}\mathcal{L}_{total}. \tag{11}$$

In terms of computational efficiency, the complexity of a single convolutional layer is expressed as

$$O\left(H \times W \times C_{in} \times C_{out} \times k^2\right). \tag{12}$$

The mathematical formulation presented in this study provides a comprehensive framework for evaluating the effectiveness of the model in classifying recipients of the Free Nutrition Food Program, both in terms of accuracy and computational efficiency. Through the formulation of supervised classification in equation (1), the construction of teacher and student models in equations (2)–(5), the application of temperature-based probability distributions in equations (6), (7), and the optimization of the joint loss function in equations (8)–(11), the student model learns not only from the actual labels but also from the probabilistic information (soft targets) generated by the teacher model. In addition, the analysis of computational complexity in equation (12) allows for a quantitative evaluation of the achieved efficiency improvements.

6. Discussion of the application of knowledge distillation in convolutional neural network algorithms

The results obtained in this study can be explained through the integration of the mathematical formulations in equations (1)–(12), the system architecture visualization in Fig. 1, and the empirical findings presented in Tables 1–5 and Fig. 2–4. Mathematically, the effectiveness of the model is explained by the combined objective function in equations (9), (10), which integrates classification loss and distillation loss. This combination enables the student model to learn not only from the actual labels but also from the probability distribution of the teacher model through the temperature mechanism in equations (6), (7). This explains why, as shown in the evaluation results and the confusion matrix in Fig. 4, the student model is able to maintain an accuracy level in the range of 80–85%, with relatively small misclassification errors in the highest-grade categories. Furthermore, the improvement in academic performance in social studies, English, and religious education, as presented in Tables 1, 5 can be explained by the associative patterns identified using the A-Priori algorithm, where support and confidence values of 0.8 indicate consistent relationships among variables. The cluster analysis in Fig. 2, 3 further reinforces the finding that the program's most significant impact occurs at the junior high school level, as evidenced by the increase in average scores before and after the intervention.

The uniqueness of the proposed method lies in the integration of the Knowledge Distillation approach in the CNN architecture for receiver classification, which is further combined with association rule mining and clustering analysis to evaluate the academic impact. Unlike conventional CNN approaches that rely solely on hard labels, this method uses soft targets to improve the model's generalization ability. Study [23] only conducted the application of knowledge distillation in image or text classification and applied the approach to social and educational data, which are still relatively underexplored in the literature. This allows for the measurement of computational efficiency through complexity analysis in Equation (12) providing a quantitative contribution to the performance evaluation, not only in terms of accuracy but also inference time and model size.

The proposed model in this study has three distinctive features. These parameters are utilized during training of the student model. In contrast to the approach in [13], where CNN and knowledge distillation were not combined, it is possible to enable the model generalization process to be faster. This an effort to help the student model learn patterns without excessive time. However, there are still several limitations to this study. The current model implementation uses only a single CNN architecture as well as one data source, making the results highly dependent on the quality of initial data and labels, as well the amount and split of test/train data. Other aspects that are dependent on the selection of parameters like learning rate, temperature and data split ratio also affect repeatability of results. The stability of the model with respect to perturbation in data distribution or other features of population has not been established extensively yet, therefore it should only be applied on conditions similar to the dataset.

This study also has some limitations; one of which was the lack of comparison with a number of other CNN architectures or classification methods including random forest, support vector machine or transformer-based models. Furthermore, a comprehensive sensitivity analysis of temperature parameter and loss weights has yet to be performed. This may be the subject of future studies by running larger experiments and more systematic hyperparameter search. Future work can extend this study, such as, increasing the scale of dataset, applying multi-teacher distillation or incorporating semi-supervised learning methods. Possible obstacles are the high-level mathematical analysis on massive data, adequate trade-off between correctness and efficiency in generalization, and constraints for gathering more comprehensive and reliable datasets. Additionally, lower-level hardware capacity and algorithm optimization are needed for real-time system development.

7. Conclusion

1. The application of knowledge distillation in CNN architectures has been shown to improve computational efficiency without compromising classification accuracy. The student model, trained using a combination of classification loss and distillation loss, maintained an accuracy rate of 80–85% with a lower model complexity than the teacher model, making it more optimal for the practical and efficient implementation of a classification system for Free Nutrition Program recipients.

2. The integration of classification methods with association and cluster analysis provides a more comprehensive evaluation of the program's impact. Association rule mining results demonstrate consistent patterns of relationships between academic variables, while cluster analysis indicates that the program's impact is most significant at the junior high school level. This demonstrates that the proposed approach is not only effective in classifying recipients but also supports data-driven analysis for more targeted policy decision-making.

Conflict of interest

The authors declare that they have no conflicts of interest related to this study, whether financial, personal, authorial, or otherwise, that could have influenced the study and the results presented in this paper.

Financing

The study was performed without financial support.

Data availability

Data will be made available on reasonable request.

Use of artificial intelligence

The authors confirm that they did not use artificial intelligence technologies when creating the current work.

Acknowledgments

I would like to thank to an honorable of Ministry of Higher Education, Science, and Technology of the Republic of Indonesia.

Authors' contributions

Relita Buaton: Conceptualization, Methodology; **Mesra Betty Yel**: Validation, Formal analysis, Investigation; **Novriyenni:** Resources, Data curation; **Anton Sihombing**: Writing – original draft, Writing – review & editing, visualization; **Ida Ria Royentina Sidabukke**: Project administration.

References

1. Arnita, A., Marpaung, F., Koemadji, Z. A., Hidayat, M., Widianto, A., Aulia, F. (2023). Selection of Food Identification System Features Using Convolutional Neural Network (CNN) Method. Scientific Journal of Informatics, 10 (2), 205–216. https://doi.org/10.15294/sji.v10i2.44059
2. Greco, A., Saggese, A., Vento, M., Vigilante, V. (2021). Effective training of convolutional neural networks for age estimation based on knowledge distillation. Neural Computing and Applications, 34 (24), 21449–21464. https://doi.org/10.1007/s00521-021-05981-0
3. Yang, H., Zhang, Y., Yin, C., Ding, W. (2022). Ultra-lightweight CNN design based on neural architecture search and knowledge distillation: A novel method to build the automatic recognition model of space target ISAR images. Defence Technology, 18 (6), 1073–1095. https://doi.org/10.1016/j.dt.2021.04.014
4. Li, Y., Luo, J., Zhang, J. (2022). Classification of Alzheimer's disease in MRI images using knowledge distillation framework: an investigation. International Journal of Computer Assisted Radiology and Surgery, 17 (7), 1235–1243. https://doi.org/10.1007/s11548-022-02661-9
5. Song, H., Yuan, Y., Ouyang, Z., Yang, Y., Xiang, H. (2024). Efficient knowledge distillation for hybrid models: A vision transformer-convolutional neural network to convolutional neural network approach for classifying remote sensing images. IET Cyber-Systems and Robotics, 6 (3). https://doi.org/10.1049/csy2.12120

6. Cho, J., Lee, M. (2019). Building a Compact Convolutional Neural Network for Embedded Intelligent Sensor Systems Using Group Sparsity and Knowledge Distillation. Sensors, 19 (19), 4307. https://doi.org/10.3390/s19194307
7. Ding, Z., Yang, C., Hu, B., Guo, M., Li, J., Wang, M. et al. (2024). Lightweight CNN combined with knowledge distillation for the accurate determination of black tea fermentation degree. Food Research International, 194, 114929. https://doi.org/10.1016/j.foodres.2024.114929
8. Ullah, H., Munir, A. (2023). A 3DCNN-Based Knowledge Distillation Framework for Human Activity Recognition. Journal of Imaging, 9 (4), 82. https://doi.org/10.3390/jimaging9040082
9. Ji, M., Peng, G., Li, S., Cheng, F., Chen, Z., Li, Z., Du, H. (2022). A neural network compression method based on knowledge-distillation and parameter quantization for the bearing fault diagnosis. Applied Soft Computing, 127, 109331. https://doi.org/10.1016/j.asoc.2022.109331
10. MohiEldeen Alabbasy, F., Abohamama, A. S., Alrahmawy, M. F. (2023). Compressing medical deep neural network models for edge devices using knowledge distillation. Journal of King Saud University – Computer and Information Sciences, 35 (7), 101616. https://doi.org/10.1016/j.jksuci.2023.101616
11. Hinton, G., Vinyals, O., Dean, J. (2015). Distilling The Knowledge In A Neural Network. arXiv. https://doi.org/10.48550/arXiv.1503.02531
12. Aghli, N., Ribeiro, E. (2021). Combining Weight Pruning and Knowledge Distillation For CNN Compression. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 3185–3192. https://doi.org/10.1109/cvprw53098.2021.00356
13. Mazlan, A. B., Ng, Y. H., Tan, C. K. (2022). A Fast Indoor Positioning Using a Knowledge-Distilled Convolutional Neural Network (KD-CNN). IEEE Access, 10, 65326–65338. https://doi.org/10.1109/access.2022.3183113
14. Chen, W., Gao, L., Li, X., Shen, W. (2022). Lightweight convolutional neural network with knowledge distillation for cervical cells classification. Biomedical Signal Processing and Control, 71, 103177. https://doi.org/10.1016/j.bspc.2021.103177
15. Song, H., Wei, C., Yong, Z. (2023). Efficient knowledge distillation for remote sensing image classification: a CNN-based approach. International Journal of Web Information Systems, 20 (2), 129–158. https://doi.org/10.1108/ijwis-10-2023-0192
16. Jomaa, L. H., McDonnell, E., Probart, C. (2011). School feeding programs in developing countries: impacts on children's health and educational outcomes. Nutrition Reviews, 69 (2), 83–98. https://doi.org/10.1111/j.1753-4887.2010.00369.x
17. Mhurchu, C. N., Gorton, D., Turley, M., Jiang, Y., Michie, J., Maddison, R., Hattie, J. (2012). Effects of a free school breakfast programme on children's attendance, academic achievement and short-term hunger: results from a stepped-wedge, cluster randomised controlled trial. Journal of Epidemiology and Community Health, 67 (3), 257–264. https://doi.org/10.1136/jech-2012-201540
18. Raveenthiranathan, L., Ramanarayanan, V., Thankappan, K. (2024). Impact of free school lunch program on nutritional status and academic outcomes among school children in India: A systematic review. BMJ Open, 14 (7), e080100. https://doi.org/10.1136/bmjopen-2023-080100
19. Hamdan, F., Al-Jarrah, F. (2024). Nutritional Health and Its Impact on Students' Academic Achievement. Journal of Educational and Social Research, 14 (6), 449. https://doi.org/10.36941/jesr-2024-0185
20. Salih, M. S., Pasha, S. A. (2024). Utilizing nutritional and lifestyle data for predicting student academic performance: a machine learning approach. Science Journal of University of Zakho, 12 (3), 356–360. https://doi.org/10.25271/sjuoz.2024.12.3.1288
21. Zarlis, M., Oktavia, T., Buaton, R., Ernawan, F., Andrian, K. (2023). Minimizing the Number of Stunting Prevalence Using the Euclid Algorithm Clustering Approach. 2023 International Conference of Computer Science and Information Technology (ICOSNIKOM), 1–7. https://doi.org/10.1109/icosnikom60230.2023.10364489
22. Kirk, D., Kok, E., Tufano, M., Tekinerdogan, B., Feskens, E. J. M., Camps, G. (2022). Machine Learning in Nutrition Research. Advances in Nutrition, 13 (6), 2573–2589. https://doi.org/10.1093/advances/nmac103
23. Umirzakova, S., Abdullaev, M., Mardieva, S., Latipova, N., Muksimova, S. (2024). Simplified Knowledge Distillation for Deep Neural Networks Bridging the Performance Gap with a Novel Teacher–Student Architecture. Electronics, 13 (22), 4530. https://doi.org/10.3390/electronics13224530