ESOGÜ Müh. Mim. Fak. Dergisi 2023, 31(1), 481-490 J ESOGU Eng. Arch. Fac. 2023, 31(1), 481-490

A NEW MODEL ON BENTHIC FORAMINIFER IMAGE CLASSIFICATION AND DEFINITIONS BASED ON CONVENTIONAL NEURAL NETWORK (CNN)

Kübra YAYAN[1*], Uğur YAYAN[2]

[1] Eskişehir Osmangazi Üniversitesi, Mühendislik Mimarlık Fakültesi, Jeoloji Mühendisliği Bölümü, Eskişehir, Turkey, ORCID No : http://orcid.org/0000-0001-7003-6437

[2] Eskişehir Osmangazi Üniversitesi, Mühendislik Mimarlık Fakültesi, Yazılım Mühendisliği Bölümü, Eskişehir, Turkey, ORCID No : http://orcid.org/0000-0003-1394-5209

Keywords

Geology, Benthic Foraminifera, Classification, Deep Learning, Convolutional Neural Networks

Abstract

Fossil studies are of great importance in observing the changes in living species over time, making inferences using information from observed species, and understanding the evolving structure of our world. However, the examination and interpretation of fossil specimens is a complex and lengthy process. Artificial intelligence (AI) studies are now being applied to this field to facilitate the working methods of paleontologists. By aiding the detection and image classification of fossil specimens, computers simplify this process in comparison to manual classification and reduce foreign dependency on fossil assemblages that paleontologists are not experts in. To achieve this, images of 9 benthic foraminiferal species were used: Bulimina tenuata, B. pagoda, Bolivina argentea, Bo. seminuda, Bo. spissa, Bo. subadvena, Epistominella smithi, Trifarina bradyi, and Takanayanagia delicate from the MD022508 benthic foraminifera dataset of the publicly available Endless Forams, alongside ordinary non-foraminiferal photographs. A single study available on the same dataset presents a study that uses deep convolutional neural networks for image classification and description of benthic foraminifera with ready-made models from the literature. In this study, a new method has been developed that achieves high accuracy rates with new models on the same dataset. At least 70% accuracy rates were achieved in the test results of the system trained with this method. This study, which reached high accuracy rates with a new method, presents a successful development for the branch of paleontology in the use of AI in microfossil identification.

BENTİK FORAMİNİFER GÖRÜNTÜ SINIFLAMASI VE TANIMLAMALARINDA EVRİŞİMLİ SİNİR AĞI (CNN) TABANLI YENİ BİR MODEL

Anahtar Kelimeler

Jeoloji, Bentik Foraminifer, Sınıflandırma, Derin Öğrenme, Evrişimli Sinir Ağları

Öz

Canlı türlerinin yıllar içindeki değişimini gözlemlemek, gözlemlenen türlerin sağladığı bilgilerden yararlanarak çıkarımlarda bulunmak ve içinde yaşadığımız dünyanın yıllar içinde gelişen ve değişen yapısını anlamak için fosil çalışmaları büyük önem taşımaktadır. Ancak fosil örneklerinin incelenmesi ve yorumlanması karmaşık ve uzun bir süreçtir. Paleontologların çalışma yöntemlerini kolaylaştırmak için yapay zeka çalışmaları bu alana uygulanmaya başlandı. Fosil örneklerinin bilgisayar yardımıyla tespiti ve görüntü sınıflandırılması, bu işlemi manuel sınıflandırma işlemlerine kıyasla mümkün olduğunca basitleştirir ve paleontologların uzman olmadığı fosil toplulukları için dışa bağımlılığı azaltır. Bunu başarmak için, seçilen bir veri setinden Halka açık Sonsuz Foram MD022508 bentik foraminifer veri kümesinden Bulimina tenuata, B. pagoda, Bolivina argentea, Bo seminuda, Bo. spissa, Bo. subadvena, Epistominella smithi, Trifarina bradyi ve Takanayanagia gibi 9 bentik foraminifer türünün görüntüleri ve sıradan foraminifer olmayan fotoğraflar kullanıldı. Aynı veri seti üzerinde mevcut olan tek bir çalışmada, derin evrişimli sinir ağları kullanılarak bentik foraminiferlerin görüntü sınıflandırılması ve tanımlamaları için geliştirilen ve literatürdeki hazır modellerle çalışılmış bir çalışma mevcuttur. Bu çalışmada, aynı veri seti üzerinde

* Sorumlu yazar; e-posta : kokur@ogu.edu.tr

geliştirilmiş yeni modellerle yüksek doğruluğa ulaşan yeni bir yöntem sunulmaktadır. Bu yöntemle eğitilen sistemin test sonuçlarında en az %70 doğruluk oranlarına ulaşılmıştır. Yeni bir yöntemle yüksek doğruluk oranlarına ulaşan bu çalışma, mikrofosil tanımlamada yapay zeka kullanımında paleontoloji dalı için başarılı bir gelişme oluşturmuştur.

Araştırma Makalesi
Başvuru Tarihi : 02.04.2022
Kabul Tarihi : 15.12.2022
https://doi.org/10.31796/ogummf.1096951

Research Article
Submission Date : 02.04.2022
Accepted Date : 15.12.2022

1. Introduction

Microfossils are generally small-sized fossils whose distinctive features are best studied with a microscope. They may contain a heterogeneous group of organism fossils of microscopic size, such as foraminifera, ostracoda, and radiolaria. Among these, foraminifers are a group of microorganisms that have been extensively studied worldwide. They are characterized by flowing granular ectoplasm, often supported by a so-called "test" shell. These microorganisms have a wide geographical distribution under the benthic and pelagic conditions of marine environments, with rich genera and species throughout the Phanerozoic era, and are single-celled animal creatures from the Cambrian to the present day (Platon and Gupta, 2001; Sakınç, 2008). Most of the niches of foraminifera are marine and stenohaline, although certain groups can survive under hypersaline conditions. Some species prefer low saline waters and are often found in brackish lagoons and estuaries (Saraswati and Srinivasan 2015). About 50,000 species of this group have been described, of which more than 4,000 species live today (Pawlowski, Esling, Lejzerowicz, Cedhagen, and Wilding, 2014).

Their shells are mostly calcareous, but a few may be pseudo-chitinous, agglutinated, or rarely siliceous. Their size varies between about 0.1 mm and 10 cm. Species and genera are distinguished by their morphological features.

Identifying among tens of thousands of genera and species in the literature requires both experience and dominance of the literature, and a highly sensitive study in order to clearly extract each feature of the fossil. Many paleontologists therefore specialize in only one group or one period. As a result of the existence and coexistence of many fossil groups, paleontologists feel obliged to seek support from each other for the identification of each group, or they leave the fossil only by describing it on the basis of its branch or family. Considering that each sub-fossil group has an expert on the foraminifers of each geological time, and even each foraminiferal genus, it is clearly seen that the discipline of foraminifera is shaped by the dominant success of individual experiences. For this reason, the research of a paleontologist working anywhere can sometimes depend on the knowledge of a paleontologist living in a different part of the world. One of the biggest potential benefits of this project is that it has the potential to serve as the heir of the experiences of valuable scientists, independent of the human factor, sourced from the common data accumulation for the field of paleontology, which is shaped by personal skills and experiences. Until recently, paleontological knowledge and experiences have been transferred to the digital environment in a limited way around the world. The directing of artificial intelligence applications, which are already very advanced, to this field is limited (Parab and Mehendale, 2021; Putzu, Caocci and Di Ruberto, 2014; Pan et al., 2020; Liao, Li and Luo, 2016; Zeggada, Benbraika, Melgani and Mokhtari, 2018; Wu and Prasad, 2017 etc.). It is thought that these developments will reduce the inaccessibility and/or foreign dependency on paleontology studies, and contribute to these studies with rapid support by using developing technology methods since it is an up-to-date idea. By collecting the data in the literature and developing correct detection methods, objective, measurable and developable microfossil identification studies can be achieved with artificial intelligence studies. Artificial intelligence is in a very good place at the moment in image identification. These studies have also begun on fossils (Zhong, Ge, Kanakiya, Marchitto and Lobaton, 2017; Ge, Zhong, Kanakiya, Mitra, Marchitto and Lobaton, 2017; Xu, Dai, Wang, Li and Wang, 2020; Mitra et al., 2019; Gutiérrez, Nouboud, Chalifour, and Voisin, 2018; Johansen and Sørensen, 2020; Hu, Limaye and Lu, 2020; Carvalho et al., 2020). It is developing more and more every day.

The aims of the paper are to reduce foreign dependency, save time, and create more reliable, accessible and easy-to-use software for researchers studying different time periods and different fossil groups.

In this study, experiments on the image classification of fossil specimens are described. The classification process was carried out using an algorithm trained with data obtained from segmented fossil views as training items in deep learning methods (Convolutional Neural Network). Physical properties, such as the number of chambers, dimensions, openings, and shape of the fossil specimen, are essential parameters for classification. A dataset containing different images of benthic foraminifers is required. The MD022508 benthic foraminifera dataset, publicly available in Endless Forams (Hsiang et al. 2019), was selected to train the model. The dataset used in this project is the Endless Forams image set, which includes 13 different benthic foraminifera species and also planktic foraminiferas, particles, and fragments (Marchant, Tetard, Pratiwi, Adebayo, and Garidel-Thoron, 2020). From this dataset, nine benthic species and non-foraminifer images were selected to train and test the model. These species are *Bulimina tenuata*, *B. pagoda*, *Bolivina argentea*, *Bo. seminuda, Bo. spissa*, *Bo. subadvena*, *Epistominella smithi*, *Trifarina bradyi*, and *Takanayanagia delicata*. The data of the selected species are divided into three different groups: 10% testing, 20% validation, and 70% training. As a result of the ongoing training process, when the trained system is tested with a new fossil sample, it aims to determine which genus the sample belongs to. With the proposed method, at least 70% accuracy rates were achieved in the test results of the trained system. This study, which achieved high accuracy rates with a new method, has created a successful development for the branch of paleontology in the use of artificial intelligence in microfossil identification.

In the following parts of the article, the method of the study is explained after previous studies in the literature are mentioned. This section explains the stages of preparing the dataset, setting up the model architecture, and the interface of the systems. The results section includes the convolutional neural network, classification, and conclusions.

2. Literature

Although there have been several online databases, presentations (Loeblich and Tappan, 1994; Foraminifera.eu; WoRMS Editorial Board, 2023), and a few software studies on foraminifera (Görmüş and Meriç, 2012; Deveciler and Akiska, 2018) in the literature, the description of foraminifera using the CNN method has not been well-documented. In fact, neural networks can be used in many areas, including geology for classifying fossil specimens, fossil segmentation and detection, and more (Zhong et al. 2017; Ge et al. 2017; Xu et al. 2020; Mitra et al. 2019; Gutiérrez et al. 2018; Johansen and Sørensen, 2020; Hu et al. 2020; Carvalho et al. 2020).

According to the study, pictures of six different species of foraminifera were taken with a microscope under 16 different controllable LED lights. A three-dimensional image was obtained by taking the maximum, minimum, and average values of the 16 images photographed. The combined images were used as input in a CNN network previously trained using a model formed from the combination of VGG16, ResNet50, and Inception V3 algorithms. For the classification process, the SVM classification method was used to conclude the process. A similar study (Ge et al., 2017) observed images of six different foraminifera species recorded under 16 different light microscopes for foraminifer segmentation. The main lines were created with the developed edge detection algorithm. After passing the views through the edge detection process, 128 deep feature maps were combined using the CNN method, and segmented images were obtained. The described method for recognizing paleontological paintings involves scaling the images to grayscale, adjusting brightness and contrast, and sharpening the images. Scale-invariant Feature Transform (SIFT) is used to find feature points and feature vectors for the pre-prepared images. K-means clustering is then used to find the distance between vector sets and feature points, and new centers are obtained according to the distribution of objects. For the nonlinear binary classification of feature vectors completed with clustering, classification is performed using SVM (Xu et al., 2020). Another study used 7 different species for the recognition of planktic foraminifers, and the images were used in the training of the CNN structure. The image was then estimated with ResNet50 and Vgg16 (Mitra et al., 2019). The article also describes a segmentation technique developed for use in diatom classification. The segmentation technique involves finding two thresholds on either side of the dominant pile using both the Rosin method and the approximate determination of the cumulative distribution function with the sigmoid function. The histogram consists of the cumulative distribution function and the sigmoid cumulative distribution function. The points where the vertical, two parallel lines drawn from the start and end points of the SCDF's vertex cut the CDF were determined as the threshold values. Thus, the value between the two thresholds is 0, and the values outside this threshold are 255. Afterwards, all connected components in the image are labeled, and the largest is determined. Then, the size of each component is normalized to the area of the largest convex shell. The minimum Euclidean distance between the largest component and the others is then calculated, and diatom reconstruction is done with the obtained data, thus removing diatoms from the background (Gutiérrez et al., 2018). In a similar study, smaller images of each sample were extracted using a simple but effective object detection scheme based on Gaussian filtering, grayscale thresholding, binary masking, and connected components. The first pass of Gaussian filtering, grayscale thresholding, and binary masking is set to remove the metallic border present in each image. Using a pre-trained VGG16 model, feature vectors were extracted from each of the foraminifer and sediment images in the dataset, and stochastic estimation was made with Monte Carlo Dropout (Johansen and Sørensen, 2020). The study also describes Dishti, a developed three-dimensional segmentation software that combines computer tomography data to create three-dimensional visuals and enables their segmentation (Hu et al., 2020). The tomography data is reconstructed with software called Mango. Another study on this subject mentions the identification process where microfossils are found directly with a pipeline developed. The developed system starts with the digitization of carbonic rocks.

Afterwards, the images obtained with microCT are respectively; quarter, half and full scale. The system was trained using the UNET model associated with ResNet34 and segmentation was performed (Carvalho et al., 2020).

3. Methods

Obtaining a reliable and open access dataset is the most important step in the methodology of this study. Therefore, the Endless Foram dataset was chosen for this purpose, and tests and validations were performed on selected species from this dataset. Convolutional Neural Networks were used to train the model, and in order to achieve high precision in the classification process, the CNN Model was selected and the training process was initiated. The paper has been prepared in accordance with research and publication ethics.

3.1. Dataset Preparation

The dataset is an essential part of training the neural network model. It needs to be properly split and labeled. For this purpose, the Endless Forams dataset was used, which meets the requirements for training the model with thousands of benthic and planktonic foraminifera images. The foraminifera views in the dataset were bright images, and a binary segmentation mask was applied to reduce the effect of brightness (Marchant et al. 2020). From this dataset, 9 species were selected to train and test the model, namely *Bulimina tenuata*, *B. pagoda*, *Bolivina argentea*, *Bo seminuda*, *Bo. spissa*, *Bo. subadvena*, *Epistominella smithi*, *Trifarina bradyi*, and *Takanayanagia delicata* as shown in Figure 1.

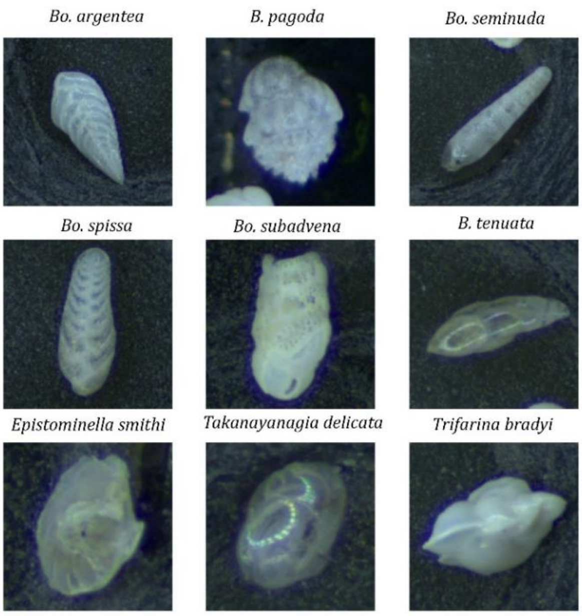

Figure1. Views of The Training Species

The data for the selected species was divided into 3 different groups: 10% for testing, 20% for validation, and 70% for training. The test and validation images were separated from the total amount of images according to the ratios mentioned earlier. To create the training set, folders were created for each species in an empty folder, and each folder was named after the corresponding species. All specimens were then separated from each other based on their species and brought together in these folders. The main folder containing all the species folders was named "training". The same process was repeated for the test and validation sets. All three main folders were then collected in another folder, concluding the dataset preparation process.

3.2. Model Architecture

To train the model, a custom convolutional neural network has been developed. Although pre-trained models might provide a faster solution, custom CNNs allow for a larger degree of network customization according to the specific purpose. The edited and tagged dataset is used, and the images are resized prior to feature extraction to a target size of 200x200. A batch size of 64 is selected. To use the feature extractions obtained from the images in CNN training, the images are divided into pixels using the Image Data Generator, a Keras function. The pixel values are scaled from 0 to 1 by dividing them by 255. This is necessary to obtain the data necessary for training the CNN model. Since the classification process will involve multiple classes, the categorical classification method is used, with the labels enumerated from zero to nine. The data used for the CNN model has now been appropriately formatted for training the model.

The CNN was chosen as the neural network to be used because high precision is needed in the classification process. CNN offers much higher precision than ANN structures due to the layers and filters it contains. In the created model, the first block is formed by a 3x3 convolutional layer. The following layer is a 2x2 pooling layer, thus forming the first block. The Convolution layer contains rectified linear (ReLU) activation. As a result of the tests, it has been observed that when the number of blocks in the model is below four, an underfitting problem occurs. Therefore, three blocks of the same structure were added following the first block. The data coming out of the neural network formed between the blocks is smoothed, and the overfitting situation is aimed to be reduced by applying dropout. Following the dropout layer, a density layer of 512 units and ReLU activation is placed. The last density layer has softmax activation and 9 units. Since 9 classes are used in the CNN model, the final density layer consists of 9 units. The representation of the model architecture is shown in Figure 2.

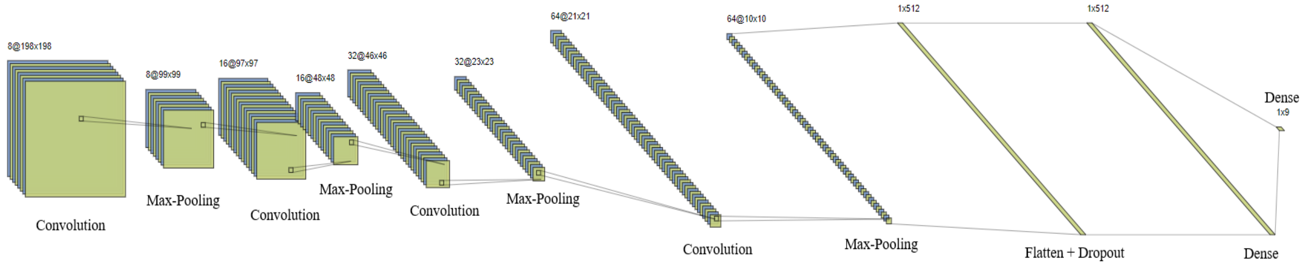

Figure 2. The Representation of The CNN Model

The training of the system was initiated using the created CNN model. CNNs are trained using categorical cross-entropy loss, which is utilized due to the implementation of the categorical classification technique as the classification mode. To update the parameters of the network, Adam (Adaptive Moment) optimization is employed. The tables below illustrate the other models developed.

3.2. Interface of The System

To simplify the process of finding the optimal system, an interface was developed. The interface allows the user to change specific hyperparameters to train the model. The interface consists of parameters such as the number of epochs, steps per epoch, batch size, image size, and the percentage of the test split of the image set. The user can input the desired parameter values via the keyboard. Upon clicking the start button, the training process of the defined model begins. The interface is shown in Figure 3.

Figure 3. Interface of The Model

4. Results

To find the optimal CNN model, many training sessions were carried out using the Endless Forams image set. This dataset contains the following species: *B. tenuata*, *Bo. argentea*, *B. pagoda*, *Bo. seminuda*, *Bo. spissa*, *Bo. subadvena*, *E. smithi*, *T. bradyi*, and *T. delicata*. The number of views for each species used to train the model are as follows: *B. tenuata* (357 views), *Bo. argentea* (236 views), *B. pagoda* (44 views), *Bo. seminuda* (778 views), *Bo. spissa* (255 views), *Bo. subadvena* (180 views), *Epistominella smithi* (92 views), *Trifarina bradyi* (41 views), and *Takanayanagia delicata* (365 views), as shown in Table 1.

Table 1: Fossil Sample Numbers and Percentage

Fossils	**Number of views**	**Number of views (%)**
B. tenuata	357	15,2
Bo. argentea	236	10,05
B. pagoda	44	1,87
Bo. seminuda	778	33,13
Bo. spissa	255	10,86
Bo. subadvena	180	7,67
Epistominella smithi	92	3,92
Trifarina bradyi	41	1,75
Takanayanagia delicata	365	15,55
Total	2348	100,0

The training and testing of the system were conducted using Tensorflow 2.5.0 with GPU enabled and Python 3.9. The computer used for training runs on Windows 10 with an AMD Ryzen 5 3600 CPU, NVIDIA RTX2060 GPU, and 16 GB RAM.

4.1. Convolutional Neural Network

To achieve the best accuracy results, the CNN model was tested multiple times as shown in Table 2. To automate the testing process, a Python loop was written to increase the specific parameter at the end of the training. The convolutional layers in the following sentences are denoted by 16C3, where 16 represents the number of filters and 3 represents the 3x3 kernel size. The max pooling layer is denoted by P2, where 2 is the size of the kernel, which is 2x2. First of all, the number of blocks is handled. The testing started with [16C3-P2]-128-9 CNN, and at every step of the test, the feature extraction blocks were increased and the number of filters doubled at every new block. The best results were obtained with [8C3-P2]-[16C3-P2]-[32C3-P2]-[64C3-P2]-1024-9, which means using 4 blocks for the convolutional layer. Other results encountered the overfitting problem more than the obtained model. From the initial test, the training accuracy of the system was obtained as 0.65 and the validation accuracy of the system was obtained as 0.52.

After conducting a parametric analysis, it can be observed that the best results were obtained when initially using 8 filters for the first convolutional layer. These results are shown in Table 3. Once the number of blocks and filters were determined, the units of the dense layer were adjusted. As shown in Table 4, the highest training accuracy and validation accuracy, 94% and 75%, respectively, were obtained when the number of units was set to 1024. To reduce overfitting, a dropout layer was added to the CNN model. However, as shown in Table 5, adding a dropout layer led to overfitting during the model training. To overcome this situation, the extra dropout layers were removed, and only one dropout layer was left before the output dense layer. The results indicate that adding dropout does not fit well with this neural network model. After conducting these experiments, it is clear from the tables that the optimum CNN structure for this work is as follows: [8C3-P2]-[16C3-P2]-[32C3-P2]-[64C3-P2]-1024-9.

Table 2. Number of Convolutional Layers and Max Pool Layers Comparison Table

Convolutional Layers	**Epoch**	**Step**	**Batch size**	**Activation**	**Training Max. Acc.**	**Validation Max. Acc.**
Single Block	100	10	64	Softmax	0.65	0.52
Two Blocks	100	10	64	Softmax	0.67	0.61
Three Blocks	100	10	64	Softmax	0.70	0.65
Four blocks	100	10	64	Softmax	0.87	0.70
Five Blocks	100	10	64	Softmax	0.82	0.67

Table 3. Number of Filters Comparison

Filters	**Epoch**	**Step**	**Batch size**	**Activation**	**Training Max. Acc.**	**Validation Max. Acc.**
8	100	10	64	Softmax	0.93	0.73
16	100	10	64	Softmax	0.90	0.72
32	100	10	64	Softmax	0.90	0.72
64	100	10	64	Softmax	0.89	0.72

Table 4. Dense Layer Unit Comparison Table

Fully connected Layer Unit	**Epoch**	**Step**	**Batch size**	**Activation**	**Training Max. Acc.**	**Validation Max. Acc.**
16	100	10	64	Softmax	0.81	0.69
32	100	10	64	Softmax	0.83	0.72
64	100	10	64	Softmax	0.81	0.70
128	100	10	64	Softmax	0.84	0.71
256	100	10	64	Softmax	0.86	0.71
512	100	10	64	Softmax	0.90	0.75
1024	100	10	64	Softmax	0.94	0.75
2048	100	10	64	Softmax	0.91	0.73

Table 5. Dropout Rate Comparison Table

Dropout	Epoch	Step	Batch size	Activation	Training Max. Acc.	Validation Max. Acc.
0.2	100	10	64	Softmax	0.81	0.59
0.3	100	10	64	Softmax	0.83	0.64
0.4	100	10	64	Softmax	0.78	0.70
0.5	100	10	64	Softmax	0.95	0.76
0.6	100	10	64	Softmax	0.90	0.67

4.2. Classification Results

The CNN model was tested using test images to measure its success in the classification process, as shown in Figure 2. The softmax activation function generated probabilities, which are percentage representations of the class to which the images belong. During the prediction process, an image may belong to more than one class, resulting in more than one probability value. However, the highest probability value is accepted as the result of the class to which the image belongs.

The Endless Foram dataset images were used to run the proposed model 15 times, and the experiment was concluded. The dataset was divided into 70% for training, 20% for validation, and 10% for testing. The average test accuracy for the Endless Foram images was 77%, with a precision of 74%, recall of 77%, and F-score of 75%. The detailed evaluation metrics are given in Table 6.

Table 6. Evaluation Metrics of Proposed Method

Evaluation Metrics of Proposed Method

	Min.	Max.	Mean	Std. Deviation
Accuracy	0.75	0.79	0.77	0.017
Precision	0.70	0.78	0.74	0.021
Recall	0.75	0.79	0.77	0.017
F-Score	0.73	0.78	0.75	0.019

The minimum, maximum, mean, and standard deviation values in Table 6 represent the statistical data obtained from running the model 15 times. The test accuracy graph resulting from the 15 runs is shown in Figure 4.

The confusion matrix is a performance metric for machine learning classification that allows the visualization of the similarity between the correct class and the predicted class. The confusion matrix for the Endless Foram dataset is shown in Figure 5. It can be observed that the most accurate prediction was made during the classification of Bo. seminuda, while the most inaccurate prediction was made during the classification of the B. Pagoda species. Bo. seminuda species can be predicted with high accuracy because it has a more distinguishable physical structure and a higher number of specimens in the dataset compared to other species. In the case of the Pagoda type, the situation is slightly different.

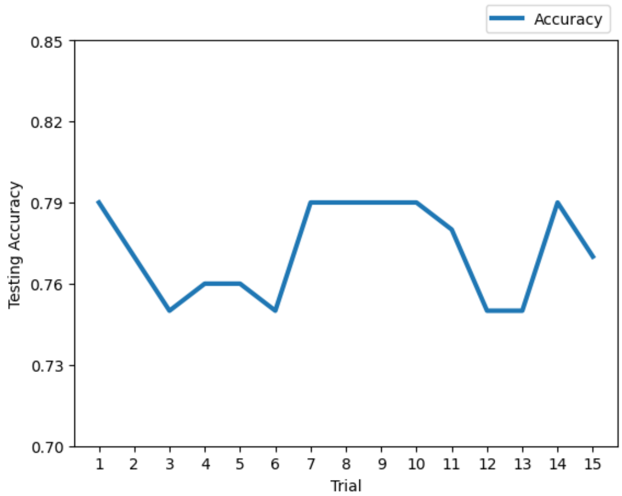

Figure 4. The Testing Accuracy Graph Obtained from Running the Proposed Method 15 Times

Although it has a different appearance from other species due to its physical structure, the scarcity of specimens in the dataset directly affects the model's learning. The abundance of data is of great importance during the training of a Neural Network for better results. Increasing the number of Pagoda specimens in the dataset can lead to more accurate predictions. The results of this study provide a basis for a potential study of adding developable code information on section directions, similar to the working methods of micropalaeontologists. Another potential benefit of this study is that it is a preliminary study for the paleontology discipline, which is shaped by personal skills and experiences, independent of the human factor, and takes its source from the common data accumulation. This study is a step forward in terms of inheriting the experience of scientists.

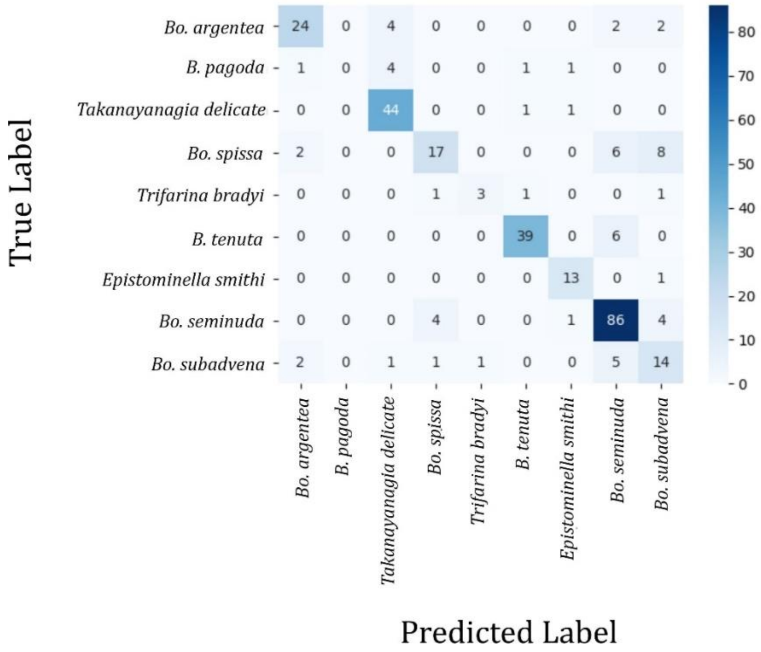

Figure 5. Confusion Matrix for Endless Foram Dataset

5. Conclusion

In this study, the CNN model developed for the classification of benthic foraminifera species and the results obtained are discussed. The main objective of this study was to create a CNN model that could classify microfossils, which are difficult to classify, more easily, quickly, and in a shorter amount of time. To achieve this, first, a dataset was created. The Endless Forams dataset was used for this purpose. The benthic foraminifera included in the Endless Forams dataset were first separated according to their species. Later, these types were divided into three main folders as training, testing, and validation folders. The training set was used to train the CNN model to be created, the validation set was used to determine the accuracy of the trained model, and the test set was used to test the operation of the created model. A lot of experimental work was done during the creation of the CNN model. As a result of the studies, it was observed that the model with four convolutional layers provided the best results. After obtaining the required model for classification, the necessary tests for the classification of the species were started. As a result of the tests, satisfactory results were obtained. Some species had higher accuracy, while others had lower accuracy. Based on the studies, it was concluded that this problem could be overcome by increasing the specimens of the species with low hit rates. When the general picture was evaluated, it could be said that the study was concluded successfully. Even though the designed model did not have enough data, it performed an acceptable classification process.

The results of this study provide a basis for potential research on the addition of developable code information on section directions, similar to the working methods of micropalaeontologists. Another significant potential benefit of this study is that it serves as a preliminary study for the field of paleontology, which is shaped by personal skills and experiences and independent of the human factor, drawing its source from common data accumulation. This study represents a significant achievement in terms of the legacy of scientific experience.

Acknowledgements

We would like to thank Engineer Metin Berke YELALDI for his help while conducting tests for this study.

Contribution of Researchers

In this study; Kübra YAYAN, identification of the problem, literature search, suitability of the data set, checking the output; Uğur YAYAN contributed to the production of suitable solution proposals for the problem, determination of working methods, explanation and control.

Conflict of Interest

No conflict of interest has been declared by the authors.

References

Carvalho, L. E., Fauth, G., Fauth, S. B., Krahl, G., Moreira, A. C., Fernandes, C. P., & Von Wangenheim, A. (2020). Automated microfossil identification and segmentation using a deep learning approach. *Marine Micropaleontology*, *158*, 101890. doi:https://doi.org/10.1016/j.marmicro.2020.101890

Deveciler, A., Akiska, S., (2018). NUMKey: An identification database program for B-forms of *Nummulites* species (foraminifera, Paleogene) for Windows and Android operating systems. Pamukkale University, Engineering Faculty Journal, 24 (6), 1179-1187. doi. 10.5505/pajes.2018.42103

Ge, Q., Zhong, B., Kanakiya, B., Mitra, R., Marchitto, T., & Lobaton, E. (2017). Coarse-to-fine foraminifera image segmentation through 3D and deep features. In *2017 IEEE Symposium series on computational intelligence (SSCI)* (pp. 1-8). IEEE. doi: 10.1109/SSCI.2017.8280982

Görmüş, M., Meriç, E. 2012. Fosil tanıma programlarının önemi: *Loftusia* örneği. SDUGEO (Online), 3 (1), 44-47 (www.geo.sdu.edu.tr), ISSN 1309-6656.

Gutiérrez Lira, E., Nouboud, F., Chalifour, A., & Voisin, Y. (2018). Image Segmentation and Object Extraction for Automatic Diatoms Classification. In *International Conference on Image and Signal Processing* (pp. 55-62). Springer, Cham. doi: https://doi.org/10.1007/978-3-319-94211-7

Hsiang, A. Y., Brombacher, A., Rillo, M. C., Mleneck-Vautravers, M. J., Conn, S., Lordsmith, S., ... & Hull, P. M. (2019). Endless Forams:> 34,000 modern planktonic foraminiferal images for taxonomic training and automated species recognition using convolutional neural networks. *Paleoceanography and Paleoclimatology*, *34*(7), 1157-1177. https://doi.org/10.1029/2019PA003612

Hu, Y., Limaye, A., & Lu, J. (2020). Three-dimensional segmentation of computed tomography data using Drishti Paint: new tools and developments. *Royal Society open science*, *7*(12), 201033. doi: https://doi.org/10.1098/rsos.201033

Johansen, T. H., & Sørensen, S. A. (2020). Towards detection and classification of microscopic foraminifera using transfer learning. *arXiv preprint arXiv:2001.04782*. doi: https://doi.org/10.48550/arXiv.2001.04782

Liao, H., Li, Y., Luo, J. (2016). Skin disease classification versus skin lesion characterization, Achieving robust diagnosis using multi-label deep neural networks. In *2016 23rd International Conference on Pattern Recognition (ICPR)* (pp. 355-360). IEEE. doi: 10.1109/ICPR.2016.7899659.

Marchant, R., Tetard, M., Pratiwi, A., Adebayo, M., & de Garidel-Thoron, T. (2020). Automated analysis of foraminifera fossil records by image classification using a convolutional neural network. *Journal of Micropalaeontology*, *39*(2), 183-202. doi: https://doi.org/10.5194/jm-39-183-2020

Mitra, R., Marchitto, T. M., Ge, Q., Zhong, B., Kanakiya, B., Cook, M. S., ... & Lobaton, E. (2019). Automated species-level identification of planktic foraminifera using convolutional neural networks, with comparison to human performance. *Marine Micropaleontology*, *147*, 16-24. doi: https://doi.org/10.1016/j.marmicro.2019.01.005

Pan, X., Jin, K., Cao, J., Liu, Z., Wu, J., You, K., ... & Ye, J. (2020). Multi-label classification of retinal lesions in diabetic retinopathy for automatic analysis of fundus fluorescein angiography based on deep learning. *Graefe's Archive for Clinical and Experimental Ophthalmology*, *258*(4), 779-785. doi: https://doi.org/10.1007/s00417-019-04575-w

Parab, M. A., Mehendale, N. D. (2021). Red blood cell classification using image processing and CNN. *SN Computer Science*, *2*(2), 1-10. Doi: https://doi.org/10.1007/s42979-021-00458-2

Pawlowski, J., Esling, P., Lejzerowicz, F., Cedhagen, T., & Wilding, T. A. (2014). Environmental monitoring through protist next-generation sequencing metabarcoding: Assessing the impact of fish farming on benthic foraminifera communities. *Molecular ecology resources*, *14*(6), 1129-1140. doi: https://doi.org/10.1111/1755-0998.12261

Platon, E., & Gupta, B. K. S. (2001). Benthic Foraminiferal Communities in Oxygen-Depleted Environments of the Louisiana Continental Shelf. *Coastal hypoxia: consequences for living resources and ecosystems*, *58*, 147-163. doi: https://doi.org/10.1029/CE058p0147

Putzu, L., Caocci, G., Di Ruberto, C. (2014). Leucocyte classification for leukaemia detection using image processing techniques. *Artificial intelligence in medicine*, *62*(3), 179-191. doi: https://doi.org/10.1016/j.artmed.2014.09.002

Sakınç M. (2008). Marmara Denizi BentikForaminiferleri: Sistemaik ve Otoekoloji [Benthic Foraminifers of the Sea of Marmara: Systemic and Autoecology. Istanbul Technical University]. İstanbul Teknik Üniversitesi. S.

Saraswati, P. K., & Srinivasan, M. S. (2015). *Micropaleontology, Principles and applications*. Springer. https://www.springer.com/gp/book/9781402059030

Wu, H., & Prasad, S. (2017). Semi-supervised deep learning using pseudo labels for hyperspectral image classification. *IEEE Transactions on Image Processing*, *27*(3), 1259-1270. doi: 10.1109/TIP.2017.2772836.

Xu, Y., Dai, Z., Wang, J., Li, Y., & Wang, H. (2020). Automatic recognition of palaeobios images under microscope based on machine learning. *IEEE Access*, *8*, 172972-172981. doi: 10.1109/ACCESS.2020.3024819

Zeggada, A., Benbraika, S., Melgani, F., Mokhtari, Z. (2018). Multilabel conditional random field classification for UAV images. *IEEE Geoscience and Remote Sensing Letters*, *15*(3), 399-403. doi: 10.1109/LGRS.2018.2790426

Zhong, B., Ge, Q., Kanakiya, B., Marchitto, R. M. T., & Lobaton, E. (2017). A comparative study of image classification algorithms for Foraminifera identification. In *2017 IEEE symposium series on computational intelligence (SSCI)* (pp. 1-8). IEEE. doi: 10.1109/SSCI.2017.8285164

Loeblich, A. R., & Tappan, H. (1994). Foraminifera of the Saharan Atlantic and Adjacent Seas: Second Edition. Science Series, Natural History Museum of Los Angeles County, 33, 1-912. Retrieved from https://foraminifera.eu/index.php?lg=en

Copernicus. (2020). Marine proxy data: Foraminifera.eu database and its applications [Conference session]. EGU General Assembly Conference Abstracts, 22, 18067. https://doi.org/10.5194/egusphere-egu2020-18067

WoRMS Editorial Board. (2023). World Register of Marine Species: Foraminifera. Retrieved April 10, 2023, from https://www.marinespecies.org/foraminifera