IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 30 October 2023, accepted 21 November 2023, date of publication 23 November 2023,
date of current version 29 November 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3336289

RESEARCH ARTICLE

Multimodal Analysis of Unbalanced Dermatological Data for Skin Cancer Recognition

PAVEL A. LYAKHOV[1,2], ULYANA A. LYAKHOVA[1], AND DIANA I. KALITA[1]
[1]Department of Mathematical Modeling, North-Caucasus Federal University, 355017 Stavropol, Russia
[2]North-Caucasus Center for Mathematical Research, North-Caucasus Federal University, 355017 Stavropol, Russia

Corresponding author: Ulyana A. Lyakhova (uljahovs@mail.ru)

This work was supported in part by the Russian Science Foundation under Project 23-71-10013, and in part by the North-Caucasus Center for Mathematical Research with the Ministry of Science and Higher Education of the Russian Federation under Agreement 075-02-2023-938.

ABSTRACT To date, skin cancer is the most commonly diagnosed form of cancer in humans and is one of the leading causes of death in cancer patients. AI technologies can match and exceed visual analysis methods in accuracy, but they carry the risk of a false negative response when a malignant pigmented lesion can be recognized as benign. Possible ways to improve accuracy and reduce the risk of false negatives are to analyze heterogeneous data, combine different preprocessing methods, and use modified loss functions to eliminate the negative impact of unbalanced dermatological data. The article proposes a multimodal neural network system with a modified cross-entropy loss function that is sensitive to unbalanced heterogeneous dermatological data. The novelty of the proposed system lies in the emerging synergy when using methods to improve the quality of intelligent systems, due to which there is a significant reduction in the number of false negative predictions and an increase in the accuracy of skin cancer recognition. Preliminary cleaning of hair structures on visual data, as well as parallel analysis of heterogeneous dermatological data using a multimodal neural network system sensitive to unbalanced data, were used as methods to improve accuracy. The recognition accuracy for 10 diagnostic categories for the proposed intelligent system was 85.20%. The introduction of weighting factors made it possible to reduce the number of false negative forecasts, as well as increase the accuracy by 1.99-4.28 percentage points compared to the original multimodal systems.

INDEX TERMS Artificial intelligence, imbalanced classification, cost-sensitive learning, multimodal neural networks, skin cancer, melanoma.

I. INTRODUCTION

To date, skin cancer is the most frequently diagnosed form of oncopathology in humans and represents a wide range of malignancies [1]. More than 40% of the total number of diagnosed cancers in the world are skin cancer [2]. The sharp increase in the incidence of skin cancer in the world is explained by chronic exposure to ultraviolet radiation (UV) [3] and the predominance of I-II skin phototypes in humans. Also, a sharp increase in diagnosed malignant pigmented lesions is due to the improvement of diagnostic methods, improved registration of data in registers, changes in the pathoanatomical criteria necessary for diagnosing skin cancer, and probable overdiagnosis [4].

Skin cancer can be divided into two types: non-melanoma and melanoma [5]. According to statistics from the World Health Organization (WHO), 325,000 new cases of melanoma were registered in 2020 [6], of which more than 17% of deaths were due to diagnosis at the last stage of oncopathology [7]. The median five-year survival rate for patients diagnosed with early-stage melanoma is about 99% [8]. In later stages, when the disease reaches the lymph nodes, the survival rate drops to 68% [9]. In the last stages, when the disease metastasizes to distant organs, the five-year survival rate is 27% [10].

Non-melanoma skin cancer (NMSC) includes basal cell carcinoma, squamous cell carcinoma, and other less common

The associate editor coordinating the review of this manuscript and approving it for publication was Rajeswari Sundararajan.

skin cancers [11]. NMSC accounts for about 1/3 of all malignant neoplasms diagnosed annually worldwide [12]. Thus, there is a need to develop balanced auxiliary diagnostic tools aimed at identifying various non-melanoma and melanoma types of malignant skin lesions, including basal cell carcinoma, squamous cell carcinoma, and others.

A significant influence on the risk of skin malignancies is exerted by such metadata factors as age, gender, localization of the pigmented lesion on the body, genetic predisposition, melanin content in the skin layers, etc. [13]. The risk of melanoma formation increases with the age of patients, as evidenced by the average age of diagnosis, which is approximately 60 years [1]. The relationship between the occurrence of malignant pigmented lesions and age becomes very clear in people over 75 years of age. The incidence of skin cancer in elderly patients doubles compared to younger patients [14]. Gender also has a significant impact on the risk of skin cancer. The incidence of melanoma in men is 1.5 times higher than in women [15]. The incidence of NMSC is also closely related to age and gender. At an early age, people of either sex show the same prevalence of any type of NMSC. However, in men older than 45 years, NMSC is diagnosed 2-3 times more often than in women [16]. Therefore, in the primary diagnosis, in addition to visual analysis, it is also necessary to take into account the complete clinical picture of each patient.

To date, the main form of skin cancer detection is a visual clinical examination using dermatoscopy [17]. Dermatoscopy is a non-invasive research method that allows one to study the diagnostically significant morphological features of pigmented skin formations [18]. The average accuracy of visual diagnosis of malignant tumors by an experienced dermatologist is 65-75% [19], [20]. This is because early diagnosis of skin cancer can be difficult due to similar morphological manifestations in benign and malignant skin lesions. The method of visual diagnostics requires extensive training and experience from a specialist in the field of dermatology [21]. If a malignancy is suspected, a histopathological examination is performed using a biopsy, which is an invasive diagnostic method. The histopathological analysis is considered the ''gold standard'' for diagnosing skin cancer. However, it is time-consuming and may be inconclusive in borderline cases. Discrepancies in diagnosis between individual pathologists can be up to 25% [22], [23].

Artificial intelligence technologies make it possible to analyze skin pigment lesions in a faster, more convenient, and more affordable way [24]. The main task of such systems is the preliminary assessment of suspicious pigmented skin lesions using high-quality histopathologically confirmed clinical images and machine learning methods [25]. However, such systems cannot replace the decisive opinion of the pathologist and dermatologist-oncologist in the diagnosis of skin cancer due to the possibility of false negative predictions [26]. Therefore, at present, the development of high-precision intelligent systems that can be used as auxiliary diagnostic tools for detecting malignant neoplasms at an early stage is becoming relevant.

One of the main problems of existing medical datasets is the asymmetric distribution of data toward the category of healthy patients [27]. The deficiency or excess of data in one or more categories can be related to the clinical characteristics of patients and disease characteristics, as well as the results of studies. As a result, a large number of negative cases of the disease are diagnosed compared with a small number of positive cases of pathologies [28]. Due to the influence of the more common category on traditional machine learning methods, the prediction results are good in the majority categories, but not accurate enough in the minority categories. There is a risk of false negative predictions, which can have potentially fatal consequences for patients.

To solve the problem of data imbalance, there are several approaches based on the transformation of training data [29], the modification of training methods [30], the development of single-class classifiers [31], and classifier ensembles [32]. The augmentation method using affine transformations increases the amount of training data due to minor changes in the color, size and shape of images [33]. However, simple operations are not enough to significantly increase the accuracy of recognition of the minority category or overcome the problem of overfitting [34]. The resampling method balances the training data and can be used as oversampling in minority categories [35], undersampling in majority categories [36], or as a combination of both methods [37]. A significant disadvantage of this method is the possibility of skipping significant diagnostic data during machine learning, as well as an increase in computational costs during data processing [38]. Another approach is to modify training methods with weighting factors, where higher losses are assigned to minority categories [39], [40]. Since the cost of classification loss is taken into account during machine learning by neural network algorithms, cost-based learning methods are the most optimal for datasets with skewed distribution [41].

Despite significant achievements in the field of introducing artificial intelligence systems for applied medical problems, the development of methods for reducing the number of false negative forecasts while simultaneously increasing the accuracy of forecasting remains relevant. The main hypothesis of the manuscript is a potential reduction in the number of false negative predictions and an increase in the accuracy of skin cancer recognition due to the resulting synergy when using various methods to improve the quality of intelligent systems. This study aims to develop a multimodal neural network skin cancer prediction system that is sensitive to imbalanced dermatological data. There are overarching criteria for the development of AI-based skin pigmentation recognition algorithms that set development and performance standards to ensure product integrity, reliability, and safety [42]. The proposed system allows for the reduction of false negative predictions through the use of a modified cross-entropy loss function and analysis of heterogeneous dermatological data

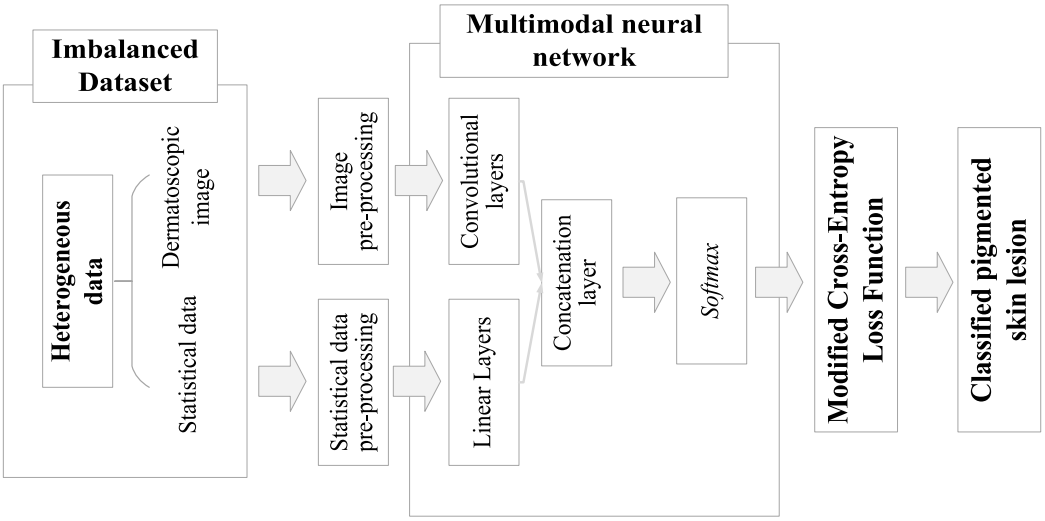

FIGURE 1. Multimodal neural network system with a modified cross-entropy loss function, sensitive to unbalanced heterogeneous dermatological data.

with the stage of preliminary cleaning of hair structures. The novelty of the proposed system lies in the simultaneous use of the method of cleaning hair structures from visual data and the further implementation of a modified cross-entropy loss function to develop a multimodal system that is sensitive to unbalanced dermatological data.

The rest of the work is structured as follows. Section II is divided into several subsections and includes a theoretical description of the proposed multimodal neural network system with a pre-cleaning of hair structures that is sensitive to unbalanced dermatological data. Section III presents the results of modeling the proposed balanced-trained multimodal neural network system for the classification of pigmented neoplasms with the stage of data preprocessing. Section IV discusses the obtained results and compares them with known works. In conclusion, the results of the work are summarized.

II. MATERIALS AND METHODS

This study proposes a multimodal neural network system with a modified cross-entropy loss function, which is shown in Figure 1. The system analyzes heterogeneous data to recognize malignant pigmented skin lesions. As heterogeneous data, information from two different modalities is used, such as dermatological images and metadata about the patient (age, gender, localization of the pigmented lesion on the body). Dermatological data undergoes a pre-processing stage to improve diagnostically significant features, as well as further transformation into the format required as input for neural network systems. For the analysis of visual data, various convolutional neural network (CNN) architectures are used, pre-trained on a set of natural ImageNet images. For metadata analysis, a neural network with a multilayer perceptron architecture, consisting of three linear layers, was chosen. The proposed system is trained using a modified cross-entropy loss function using weighting factors. Weighting coefficients are calculated in a special way for a selected database of unbalanced dermatological data.

A. PRE-PROCESSING OF HETEROGENEOUS DERMATOLOGICAL DATA

The most common types of data in the field of dermatology are visual multidimensional data and metadata. Metadata includes gender, age, as well as the localization of the pigmented neoplasm on the patient's body. Visual clinical examination of the skin is the main form of diagnosing oncopathologies. The metadata of the patient may also indicate the risk of developing malignant forms of pigmented skin lesions [43]. Therefore, there is a need for a comprehensive analysis of heterogeneous data for more accurate diagnosis [44]. The combination of visual data as well as metadata about patients makes it possible to create heterogeneous databases of dermatological information that can be used to build intelligent diagnostic and decision support systems for specialists, doctors and clinicians [45]. The use of heterogeneous information makes it possible to increase the accuracy of neural network analysis by searching for additional links between images and metadata [46].

Diagnostically significant multidimensional visual information can be distorted by noises of various natures, as well as by various physiological factors. The presence of hair structures on dermatological images violates the geometric properties of the site of the pigmented neoplasm [47]. These noise distortions can drastically change the size, shape, color, and texture of a lesion. This significantly affects the result of the analysis of auxiliary automated diagnostic systems. Hair removal from images at the preprocessing stage is an important step in the development of systems based on artificial intelligence [48]. For preliminary digital processing of visual data, a method was proposed for cleaning hair structures using morphological operations. The proposed method is presented in the paper [49] and consists of four stages. In the

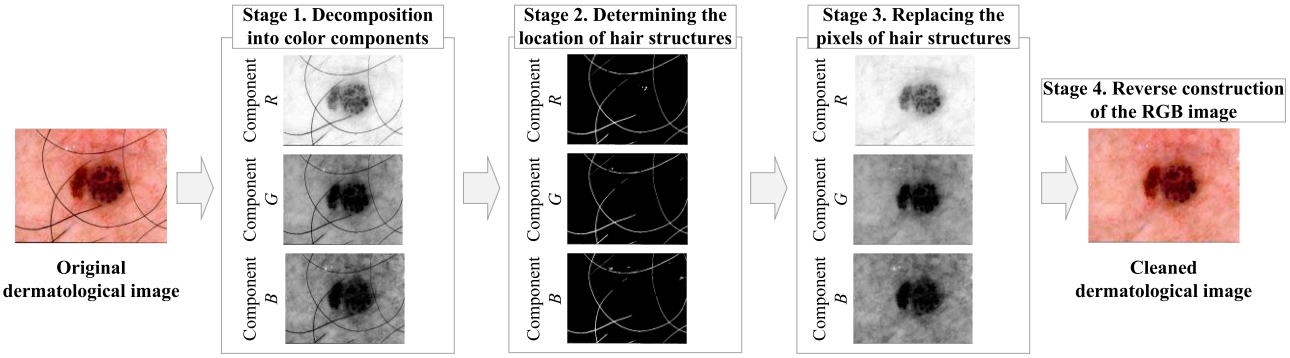

FIGURE 2. An example of the step-by-step operation of the method of pre-cleaning of hair structures on dermatological images.

first stage, the processed dermatological RGB image of the pigmented neoplasm is decomposed into color components. Further processing is performed separately for each color component. In the second stage, the location of hair structures is determined using a morphological closing operation with a given element. In the third stage, hair pixels are replaced with neighboring pixels using interpolation. In the fourth stage, the reverse construction of the dermatological RGB image from the color components is performed. The use of the proposed method for cleaning hair structures can significantly improve the recognition accuracy of the neural network system by improving the quality of visual diagnostically significant information. An example of the phased work of the used method for pre-cleaning hair structures is shown in Figure 2.

Most systems based on artificial intelligence require as input a feature vector in the form of integers [50]. Converting categorical variables to a numerical format is a necessary step for correctly calculating the correlation between them and further intelligent prediction [51]. The most common way to preprocess categorical variables is the one-hot encoding method [52]. As a result, categorical variables with multiple possible values are transformed into a new set of numeric position vectors, all elements of which are equal to zero, except for the position of the variable value in the list of all possible values [53]. The processed metadata S includes a certain number of patient factors:

$$S = \{s_1, s_2, \ldots, s_q\}; s_q \in S_q, \tag{1}$$

where S_q is the patient's metadata factor; s_q is a pointer to a specific patient variable. If S_1 indicates the localization factor of the pigmented neoplasm on the patient's body, then s_1 can take one of eight possible values, such as localization on the anterior torso, head/neck, lateral torso, lower extremity, oral/genital, palms/soles, posterior torso or upper extremity.

When processing metadata using the one-time coding method, the dimension of the input feature vector $\vec{S}$ is formed as follows:

$$dim\,\vec{S} = \sum_q \varphi_q = \sum_q |S_q|, \tag{2}$$

where φ_q is the cardinality of the metadata factor S_q, which depends on the number of all possible values of the factor.

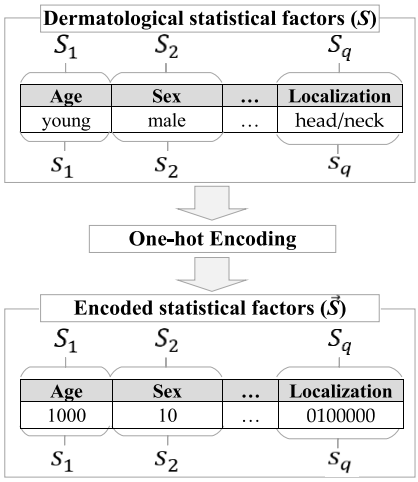

FIGURE 3. Scheme for processing dermatological metadata using the one-hot encoding method.

The scheme for processing dermatological metadata using the one-hot encoding method is shown in Figure 3.

B. MODIFICATION OF THE CROSS-ENTROPY LOSS FUNCTION USING WEIGHT COEFFICIENTS

With the development of artificial intelligence technologies and the advent of large amounts of digital information, machine learning algorithms began to strive to increase the speed and accuracy of extracting information from processed data. The main goal of machine learning algorithms is to solve the problem of optimization-minimization of structural risks, which can be represented as follows:

$$N = \min_f \frac{1}{K} \sum_{i=1}^{K} C_\theta\left(f\left(x_i\right)\right) + \Lambda M(f), \tag{3}$$

where K is the number of examples in the training set; C is the error function with variable vector θ; M is the regularization element that represents the complexity of the model; $\Lambda \geq 0$ is the balance between empirical risk and the complexity of the neural network model.

The loss function is the main part of training a neural network model and is used to adjust the weights of the neural network [54]. As a result of processing training examples by a neural network, output responses are generated that indicate

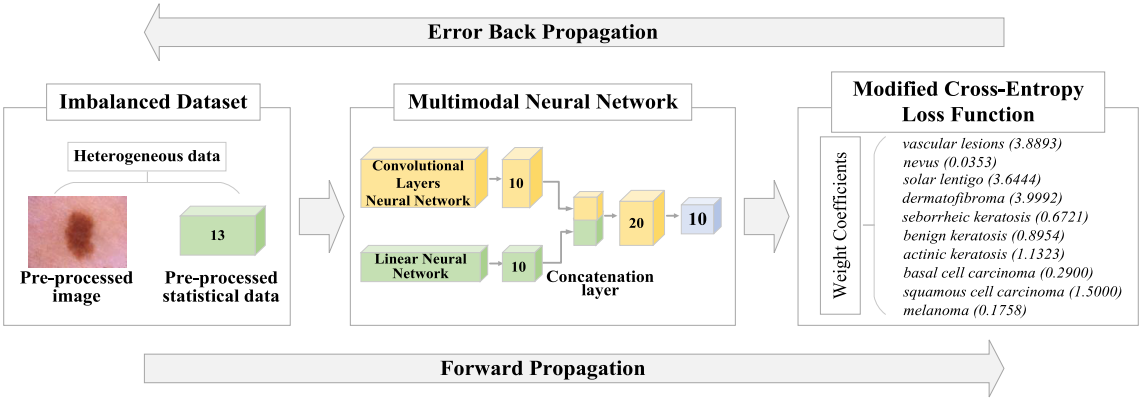

FIGURE 4. Scheme of using a modified cross-entropy loss function for training a multimodal neural network system for recognizing pigmented skin lesions.

the probability or reliability of possible categories to which the analyzed data belong [55]. The resulting probabilities are compared with the true labels. The loss function calculates a penalty for any deviation between the true label and the output of the neural network [56]. As a rule, in deep learning, the use of the logarithmic loss function is considered the most optimal [57]. To solve the problem of multiple classification, the *softmax* C_{ce} cross-entropy loss function is used, which has the form:

$$C_{ce} = -\frac{1}{K}\sum_{n=1}^{N}\sum_{i=1}^{K} l_i^n \times \log\left(h_\mu\left(x_i, n\right)\right), \quad (4)$$

where N is the number of categories in the training database; l_i^n is the true label for training case i from category n; x_i is the input of training case i; h_μ is the neural network model with weights μ.

The number of diagnosed benign cases of pigmented skin lesions in medical education databases can be twice the number of diagnosed cases of malignant neoplasms [58]. At the same time, if we consider the existing databases of medical dermatological data with a large number of diagnostic categories, it can be established that the difference between the most common and the least common category can be more than a hundred times. For example, in the public database HAM10000, the most common category is ''nevi'', which consists of 6705 examples, and the least common category is ''dermatofibroma'', which consists of 115 examples. For the category ''melanoma'' in the specified database of dermatological database contains only ''1113'' example. If the training examples contain a significant imbalance in the data, then the AI-based classifier will tend to focus on the categories with the largest number of samples [59]. The standard loss function will be successfully minimized when the neural network classifier predicts all input data as ''benign'' [60]. Thus, there may be a shift in classification efficiency towards the prevailing category [61]. To solve this problem, the most optimal is the use of unequal costs for misclassification between categories, which can be defined as a cost matrix or weighting coefficients [62].

The cost of expenses can be considered as a penalty coefficient, which is introduced when training a neural network model [63]. When developing systems for intelligent recognition of pigmented skin lesions, the penalty factor is aimed at increasing the significance of the least common classes of malignant pigmented neoplasms. As a result, there is a stronger penalty for misclassifying data from a minority sample. The neural network classifier focuses on and more carefully analyzes the data coming from this distribution. The calculation of the cost of training costs d_n is inversely proportional to the frequency of categories in the database and has the following form:

$$d_n = \frac{K}{N\sum_{i=1}^{K} p_{in}}, \quad (5)$$

where N is the number of dermatological categories; p_{in} is an indication that image i belongs to category n.

The modification of the cross-entropy loss function using weight coefficients C'_{ce} can be represented as follows:

$$C'_{ce} = -\frac{1}{K}\sum_{n=1}^{N}\sum_{i=1}^{K} d_c \times l_i^n \times \log\left(h_\mu\left(x_i, n\right)\right), \quad (6)$$

where d_n is the weighting factor for category n.

Thus, modification of the cross-entropy loss function will minimize the impact of unbalanced data and avoid bias in the classification results towards the more common category of benign skin lesions. Figure 4 shows a diagram of the application of the modified cross-entropy loss function for training a multimodal neural network system for recognizing pigmented skin lesions.

C. MULTIMODAL NEURAL NETWORK SYSTEM FOR THE ANALYSIS OF UNBALANCED DERMATOLOGICAL DATA

To date, multimodal machine learning is a promising area of research in which models are developed to analyze information from several modalities [64]. The fusion of heterogeneous data takes into account the representation of features of various modalities for a more complete analysis and allows the use of multidimensional heterogeneous information for making decisions in a neural network model [65]. The non-obvious relationship between the processed data and the results of diagnostics is extracted through an additional

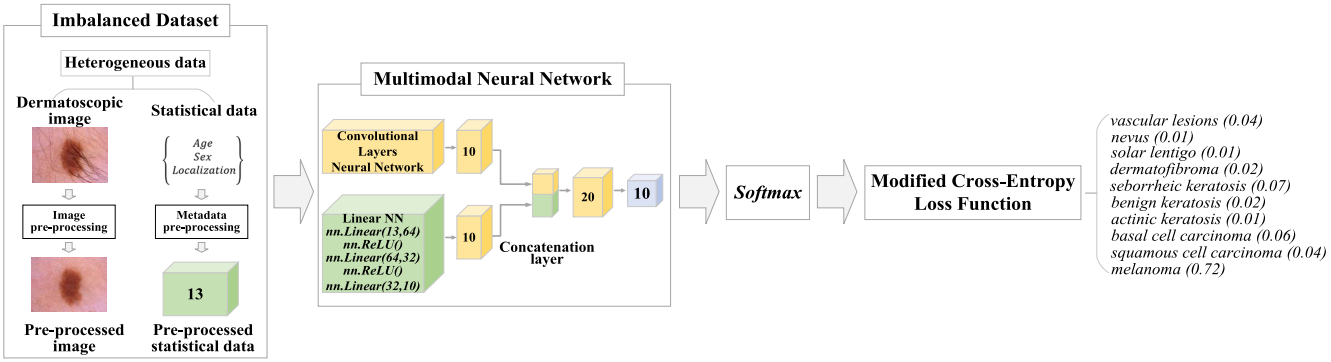

FIGURE 5. The architecture of the proposed multimodal neural network system for recognizing pigmented skin lesions with a modified cross-entropy loss function.

neural network study of information between modalities. Thus, neural networks can use additional data by integrating several modalities into a common structure [66].

CNNs are the most optimal neural network architecture for recognizing multidimensional visual data [67]. For the analysis of metadata, the most optimal is a multilayer feed-forward perceptron [68]. The proposed multimodal neural network system consists of two neural network architectures. For the analysis of visual dermatological data presented by images of pigmented skin lesions, the CNN architecture is used. A linear multilayer perceptron is used to process metadata vectors.

The input of the proposed multimodal neural network system receives pre-processed dermatological images D_{rgb}, the vector of metadata features $\vec{S}$. In a multilayer perceptron, neurons perform the summation of the received input data vector $\vec{S}$ and the bias coefficient b, forming a synaptic input. As a result of training, the weights of neurons are iteratively formed as follows:

$$v^{a+1} = w^a + \left(-r \times \frac{\partial E}{\partial w}\right), \tag{7}$$

where r is the learning rate; $\frac{\partial E}{\partial w}$ is the error gradient concerning the weights. After the signal passes through the *ReLU* activation function, the output signal of the neuron is calculated:

$$v^s = f\left(\sum\nolimits_i^n z_i v_i + b\right), \tag{8}$$

Obtaining feature maps as a result of processing D_{rgb} dermatological images is performed in parallel as follows:

$$D_f(x, y) = b + \sum\nolimits_{i=\frac{-w-1}{2}}^{\frac{w-1}{2}} \sum\nolimits_{j=\frac{-w-1}{2}}^{\frac{w-1}{2}} \sum\nolimits_{n=0}^{D-1} w_{ijn}^{(1)} \times D(x+i, y+j, n), \tag{9}$$

where D_f is the feature map of the dermatological image; $w_{ijn}^{(1)}$ is the $c \times c$ size filter factor, b is the offset factor.

On the concatenation layer, the resulting feature map D_f and the output signal v^s are combined as follows:

$$E = \sum\nolimits_i \sum\nolimits_j \sum\nolimits_n D_f w_{ijn}^{(2)} + \sum\nolimits_{i=1}^{n} v^s w_{in}^{(3)}, \tag{10}$$

where $w_{ijn}^{(2)}$ is the weight for processing feature maps of dermatological images D_f; $w_{in}^{(3)}$ is the weight for processing the output signal of the multilayer perceptron.

The last layer of the multimodal neural network system is activated through the function $softmax\ D(y \mid x, \theta) = softmax(x; \theta)$. After that, the resulting output probability distribution between categories is compared with the original correct distribution. The modified cross-entropy loss function C'_{ce} is used only in neural network systems with the specified output function. As a result, the loss function C'_{ce} indicates the distance between the output distribution and the original probability distribution. There is a gradual memorization of true vectors and a minimization of losses during training. The architecture of the proposed multimodal neural network system with a modified cross-entropy loss function is shown in Figure 5. As a result of preliminary processing of metadata, a vector of input features is formed, consisting of 13 variables. At the end of the process of processing visual data using the convolutional layers of the neural network architecture, an output layer is formed, consisting of 10 neurons corresponding to the number of recognized classes. Similarly, an output layer is formed from a linear neural network architecture, which also consists of 10 neurons corresponding to the number of recognized classes.

D. METHOD FOR STATISTICAL EVALUATION OF FALSE-NEGATIVE SKIN CANCER IN NEURAL NETWORK SYSTEMS

When developing systems for recognizing pigmented skin lesions based on artificial intelligence, it is important to monitor and minimize the risk of false negative prediction, when a malignant pigmented formation can be recognized as benign. When testing neural network systems for multiclass recognition of dermatological data, the calculation of standard statistical evaluation metrics is performed separately for each class, after which the average value is displayed [69]. For this, a standard multiclass confusion matrix is used, shown in Figure 6.

With this approach, for each class, errors in both benign and malignant categories are the same false negative (FN).

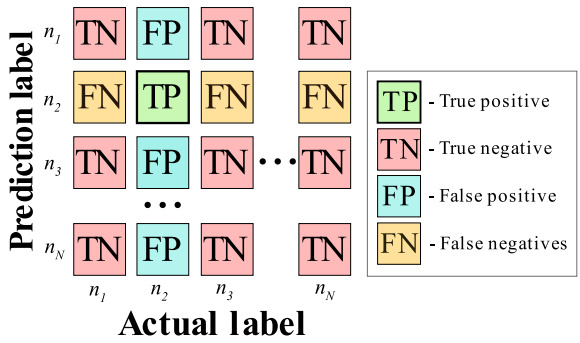

FIGURE 6. An example of a confusion matrix for class n_2 as a result of testing a multiclass neural network system.

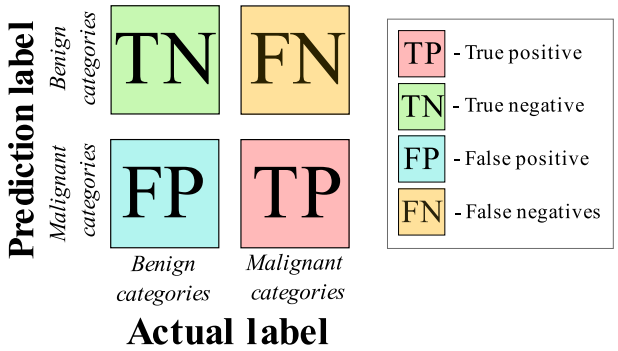

FIGURE 7. An example of a confusion matrix for benign and malignant categories as a result of testing a multiclass neural network system.

For example, when recognizing data from the ''melanoma'' class, the system response result as the ''nevus'' class, which is a benign category, so the system response result and the ''squamous cell carcinoma'' class, which is a malignant category, will be the same false negative.

The most common standard statistical metrics for evaluating intelligent systems are specificity (Sp), sensitivity (Se), $F1$, Matthew's correlation coefficient (MCC), false negative rate (FNR), false positive rate (FPR), negative predictive value (NPV) and positive predictive value (PPV). When testing intelligent medical systems in which there are categories with diseases and categories with healthy patients, standard metrics are not able to fully assess the risk of false negative prediction, which is the most significant parameter for auxiliary diagnostic systems. Therefore, in addition to the existing methods for evaluating neural network systems, it is necessary to introduce specialized metrics for assessing the quality of models.

The proposed specialized metrics for evaluating the quality of models are calculated based on the results of the standard multiclass confusion matrix. To do this, the summation of all system-predicted responses by benign and malignant categories is performed. All predicted responses of the neural network in the classes of benign categories, when the true values also belonged to benign categories, were combined into a group of true positive responses TP_{onco}. All predicted responses in the classes of benign categories, when the true values were in the malignant categories, were combined into a group of false positive responses FP_{onco}. Predicted responses in classes of malignant categories when the true values belonged to benign categories were combined into a group of false negative responses FN_{onco}. Predicted responses in classes of malignant categories when the true values were related to malignant categories were combined into a group of true negative responses TN_{onco}. As a result, the confusion matrix shown in Figure 7 is calculated.

The calculation of the proposed statistical estimates $F1_{onco}$, MCC_{onco}, Sp_{onco}, Se_{onco}, FNR_{onco}, FPR_{onco}, NPV_{onco} and PPV_{onco} is performed in the same way as for binary classification systems and can be represented as in (11), shown at the bottom of the next page.

Sp_{onco}, Se_{onco}, FNR_{onco}, FPR_{onco}, NPV_{onco} and PPV_{onco} together with standard metrics for evaluating intelligent systems will allow more reliable assessment of the risk of false negative prediction in systems of auxiliary medical dermatological diagnostics.

III. RESULTS

For practical modeling, data were selected from the open archive of the International Skin Imaging Collaboration (ISIC) [70]. The ISIC Archive is an open-source platform that contains publicly available dermatological data under a Creative Commons license. Images of pigmented skin lesions are associated with patient metadata and histopathologically confirmed diagnoses. The purpose of the archive is to provide open access to diagnostic dermatological data for training specialists in melanoma recognition methods, as well as for the development of clinical decision support systems and automated diagnostics. At the time of data selection for modeling, 41,725 dermatological images of various sizes and 41,725 associated metadata were publicly available. Moreover, each case of a pigmented skin neoplasm in the archive is assigned its own unique identification number, which eliminates the possibility of duplicates appearing in the generated databases for modeling. To select data for modeling from the ''DIAGNOSTICS OF LESIONS'' section of the ISIC archive, 10 diagnostically significant categories were identified, which included the largest volume of dermatological data. These categories were Nevus (27,878 examples), Melanoma (5,597 examples), Basal cell carcinoma (3,393 examples), Seborrheic keratosis (1,464 examples), Benign keratosis (1,099 examples), Actinic keratosis (869 examples), Squamous cell carcinoma (656 examples), Solar lentigo (270 examples), Vascular lesions (253 examples) and Dermatofibroma (246). The categories not included in the modeling database included fewer than 100 samples of pigmented skin lesions and did not allow the artificial intelligence system to be reliably trained for the classification task. For each specified category, all available dermatological images were downloaded from the ISIC archive at the time the modeling database was formed. In this case, the main part of the selected images were dermoscopic, but there were also clinical type images and close-up images.

During the processing phase, all known duplicates were searched and removed from the modeling database to improve reliability and reduce bias in deep learning models. In the first step, all selected data was manually checked for the presence of known duplicates in the database from duplicate image list [71]. All matching examples were removed from the data set and replaced with other examples from the same category from the ISIC archive. In the second step, the selected database was checked using the "ISIC Datasets Duplicate Removal Strategy" [72] script to identify and remove the remaining duplicate images. After replacing all duplicates, the modeling database went through a second stage of reconciliation with the list of duplicates and verification using the "ISIC Datasets Duplicate Removal Strategy" script to completely eliminate possible repetitions. In the modeling database, 107 examples of dermatological data matching the duplicate image list and 31 examples detected using the "ISIC Datasets Duplicate Removal Strategy" script were replaced. To avoid data leakage, it is common practice to include only one example per patient in the dataset. For each example of a pigmented skin neoplasm, the "patient_id" parameter is specified in the metadata section. At the stage of selecting data for modeling, only one example of a pigmented skin tumor was included in the database for each patient. Duplicates that appeared with a matching "patient_id" parameter were removed from the database and replaced with similar pigmented neoplasms from this category. Thus, the selected data for modeling included 41,725 dermatological images of varying sizes and qualities. Each image was associated with a set of metadata factors and an established diagnosis. The selected categories are divided into "malignant" and "benign" groups and arranged in descending order of risk and severity of the course of the disease. Actinic keratosis is an intraepithelial dysplasia of keratinocytes and is characterized as a "precancerous" skin lesion (in situ squamous cell carcinoma). Therefore, this category was assigned to the group of "malignant" pigmented skin lesions [73]. A graph of the distribution of selected dermatological images by category is shown in Figure 8.

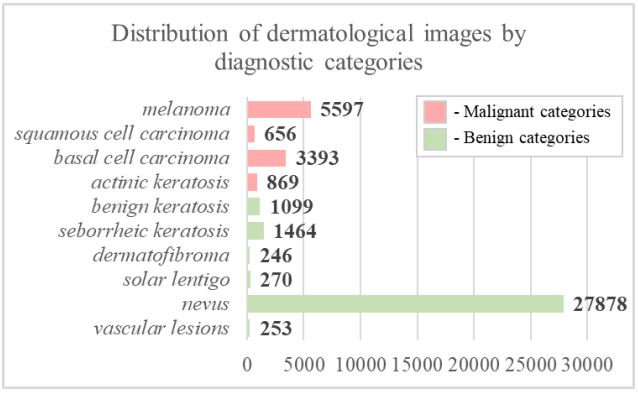

FIGURE 8. Graph of the distribution of selected dermatological images into diagnostically relevant categories.

The metadata set for each image included information about the patient's gender (male/female), age group in five-year increments, and localization of the pigmented lesion on the body (anterior torso, head/neck, lateral torso, lower limb, oral/genital area, palms/soles, hindquarters, upper limb). At the stage of preliminary processing of metadata, the "Age" variable was divided into four groups by the age classification adopted by the World Health Organization (WHO). Thus, the variability of the "Age" variable was reduced from 18 to 4 possible values. Distributions of dermatological data by various metadata factors are presented in Table 1. As a result of the analysis of metadata, it was found that the predominant number of patients refers to men and the age group up to 44 years. Also, pigmented lesions are most often localized on the posterior torso. The data obtained are highly correlated with studies on the influence of metadata factors on the risk of skin cancer [1], [13], [14], [15].

$$
\begin{aligned}
F1_{onco} &= \frac{2 * TP_{onco}}{2 * TP_{onco} + FP_{onco} + FN_{onco}};\\
MCC_{onco} &= \frac{(TP_{onco} * TN_{onco}) - (FP_{onco} * FN_{onco})}{\sqrt{(TP_{onco} + FP_{onco})(TP_{onco} + FN_{onco})(TN_{onco} + FP_{onco})(TN_{onco} + FN_{onco})}};\\
Sp_{onco} &= \frac{TN_{onco}}{TN_{onco} + FP_{onco}};\\
Se_{onco} &= \frac{TP_{onco}}{TP_{onco} + FN_{onco}};\\
FNR_{onco} &= \frac{FN_{onco}}{FN_{onco} + TP_{onco}};\\
FPR_{onco} &= \frac{FP_{onco}}{FP_{onco} + TP_{onco}};\\
NPV_{onco} &= \frac{TN_{onco}}{TN_{onco} + FN_{onco}};\\
PPV_{onco} &= \frac{TP_{onco}}{TP_{onco} + FP_{onco}}.
\end{aligned}
\tag{11}
$$

TABLE 1. Distribution of dermatological metadata in the modeling database.

Metadata factor	Value	Frequency	Percentage
Gender	female	18,431	44.17
	male	23,294	55.83
Age	≤ 44 years	17,927	42.96
	45 ≤ 59 years	10,720	25.69
	60 ≤ 74 years	7788	18.67
	≥ 75 years	5290	12.68
Localization on the body	posterior torso	17,326	41.52
	anterior torso	7440	17.83
	lower extremity	7182	17.21
	head/neck	5375	12.88
	upper extremity	3844	9.21
	palms/soles	411	0.99
	lateral torso	83	0.20
	oral/genital	64	0.15

Training and testing were carried out using the high-level programming language Python 3.11.0. All calculations were carried out on a PC with an Intel(R) Core(TM) i5-8500 processor at 3.00 GHz with 16 GB of RAM and a 64-bit Windows 10 operating system. Training of multimodal neural network systems was carried out using a graphics processing unit (GPU) based on NVIDIA GeForce GTX 1050TI video chipset. The Pytorch machine learning framework was used to model neural network systems. The NumPy, Pandas, and ScikitLearn libraries were used to process metadata. The Matplotlib library was used to visualize the data.

To model a multimodal neural network system for recognizing pigmented skin lesions, sensitive to unbalanced data, neural network architectures DenseNet_161 [74], Inception_v4 [75], ResNeXt_50 [76]. The selected convolutional architectures were pre-trained on the ImageNet natural image set. To date, the selected neural network architectures are recognized as the most productive and highly accurate compared to human capabilities [77].

At the first stage of modeling, preliminary processing of dermatological data was carried out. The preprocessing of the metadata was to create an input vector using the one-hot encoding method. Thus, it was possible to reduce the number of possible values that the metadata factors of patients can take from 28 to 14. Pre-processing of visual data consisted in applying the proposed method for removing hair structures from [49]. Examples of pre-processed dermatological images are shown in Figure 9. The second step in pre-processing the visual data was to transform the size of the input data. The main part of the selected images of pigmented skin lesions from the ISIC archive is presented in the size of 450 × 600 pixels. For the selected neural network architectures, the requirements for input visual data are 256 × 256 pixels for the DenseNet_161 [74] and ResNeXt_50 [76] architectures, 229 × 229 pixels for the Inception_v4 architecture [75]. Therefore, at the stage of pre-processing, the operation of transforming the size of the input images was applied.

In accordance with the gold standard of machine learning for further modeling, the dermatological database was divided into training data and test data in a ratio of 80 to 20 (33,380 training cases and 8,345 test cases). The training data was also divided into training and validation data by a percentage of 75 to 25 (25,035 training cases and 8,345 validation cases). Thus, all data were divided into groups "training data" in the amount of 25,035 cases, "validation data" in the amount of 8,345 cases and "test data" in the amount of 8,345 cases. The data set was divided into training, validation and test so that the ratio between the visual data in each category was proportional for both the training set and the validation and test set. For example, for the training sample, the ratio of data between the categories "nevus" and "melanoma" was 5 to 1. In the validation and test samples, the ratio of data between the categories "nevus" and "melanoma" also was 5 to 1. The data were distributed similarly for all 10 diagnostic significant categories. In this case, the distribution of data took into account both the relationship of images between diagnostic categories and the relationship between metadata. Thus, all samples equally included data from patients with different pigmented lesions, different genders, different ages, and different locations on the body. Test data did not affect the learning process of all multimodal neural network systems and was used for final testing and calculation of statistical evaluation metrics. Affine transformations such as reflection, rotation, translation, scaling, etc. were applied to the training set of visual data. Data augmentation made it possible to avoid retraining neural network models. The table 2 shows the distribution of dermatological images by category in the training, validation and test sets. The table 3 shows the distribution of dermatological metadata in the training, validation and test sets.

TABLE 2. Distribution of dermatological images in training, validation and test sets.

Category	Training set	Validation set	Test set
vascular lesions	151	51	51
nevus	16730	5574	5574
solar lentigo	162	54	54
dermatofibroma	146	50	50
seborrheic keratosis	878	293	293
benign keratosis	659	220	220
actinic keratosis	521	174	174
basal cell carcinoma	2035	679	679
squamous cell carcinoma	394	131	131
melanoma	3359	1119	1119

For the training process, preprocessed dermatological images of pigmented skin lesions were fed into the input of selected CNN from the training set. The vector of preprocessed metadata from the training sample was fed to the input of the developed multilayer neural network architecture, consisting of three linear layers and ReLu activation layers. After the multimodal signals passed through the CNN and the linear perceptron, the output feature vectors were combined on the concatenation layer. The output signal was applied to the *softmax* layer to determine the probabilistic ratio of predicted labels for 10 diagnostically significant categories. The obtained probabilities were compared with the true labels to the training data, and the error value was calculated using the modified cross-entropy loss function. Neural network

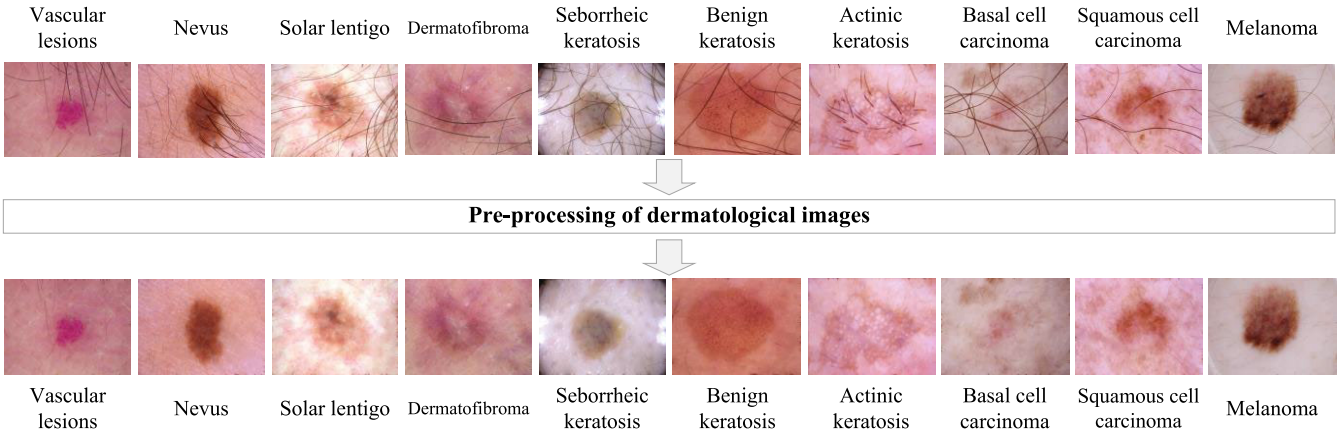

FIGURE 9. An example of pre-processed dermatological images using the hairline cleaning method.

TABLE 3. Distribution of metadata in training, validation and test sets.

Metadata factor	Value	Training set	Validation set	Test set
Gender	female	10,977	3727	3727
	male	14,060	4617	4617
Age	≤ 44 years	10,807	3560	3560
	45 ≤ 59 years	6384	2168	2168
	60 ≤ 74 years	4786	1501	1501
	≥ 75 years	3060	1115	1115
Localization on the body	posterior torso	10446	3440	3440
	anterior torso	4342	1549	1549
	lower extremity	4372	1405	1405
	head/neck	3333	1021	1021
	upper extremity	2230	807	807
	palms/soles	231	90	90
	lateral torso	47	18	18
	oral/genital	36	14	14

TABLE 4. Weight coefficients are used to modify the cross-entropy loss function in a multimodal neural network system.

№	Diagnostic category	Weight coefficient
1	Vascular lesions	3.8893
2	Nevus	0.0353
3	Solar lentigo	3.6444
4	Dermatofibroma	3.9992
5	Seborrheic keratosis	0.6721
6	Benign keratosis	0.8954
7	Actinic keratosis	1.1323
8	Basal cell carcinoma	0.2900
9	Squamous cell carcinoma	1.5000
10	Melanoma	0.1758

architectures were penalized more severely for errors in the less common categories than errors in the more common ones. As a result, there was a gradual memorization of true vectors and a minimization of losses during training. The calculated weight coefficients of each of the classes for modifying the cross-entropy loss function are presented in Table 4. A significant difference between the proposed method for modifying the cross-entropy loss function and the functionality built into PyTorch for creating a balanced data loader is the individual calculation of the weight of each category based on the number of examples in the training set.

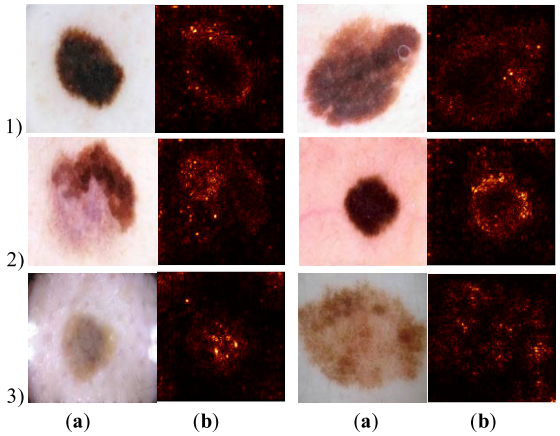

FIGURE 10. Examples of significance maps for the proposed intelligent systems that are sensitive to unbalanced data based on architecture: 1) DenseNet_161; 2) Inception_v4; 3) ResNeXt_50. Original image of a pigmented skin lesion (a). The resulting significance map as a result of intellectual analysis (b).

The size of the input data packet was 8. SGD was used as an optimizer with a standard learning rate of 0.001 and a moment of 0.9. The k-fold cross validation method was used for training. In this case, the value of k was set equal to 5. As a result of 5-fold modeling, the most accurate models were selected for further testing. Figures 18-22 from Supplementary A show graphs of the training results of multimodal neural network systems for analyzing heterogeneous dermatological data for recognizing pigmented skin lesions based on various CNN architectures. Figure 10 shows examples of significance maps for the proposed intelligent systems that are sensitive to imbalanced data based on various convolutional architectures.

Significance maps indicate in which area of the image the neural network is concentrated. The brighter and more intense the area on the significance map, the more significant the pixels in that area are for image classification. As a result of the analysis of the obtained significance maps, it can be concluded that intelligent systems based on various architectures focus on the area of pigmented skin neoplasm in the images,

TABLE 5. Results of testing multimodal neural network systems by quantitative assessment methods.

CNN architecture	Loss function weights	Accuracy, %	Loss function	F-1 score	MCC	Specificity	Sensitivity	FNR	PPV	FPR	NPV
DenseNet_161	Not used	80.92	0.2964	0.8092	0.6496	0.9788	0.8092	0.1908	0.8092	0.0212	0.9788
[74]	Used	**85.20**	**0.1294**	**0.8520**	**0.7168**	**0.9836**	**0.8520**	**0.1480**	**0.8520**	**0.0164**	**0.9836**
Inception_v4	Not used	81.34	0.2641	0.8134	0.6413	0.9793	0.8134	0.1866	0.8134	0.0207	0.9793
[75]	Used	83.63	0.1711	0.8363	0.6843	0.9818	0.8363	0.1637	0.8363	0.0182	0.9818
ResNeXt_50	Not used	82.41	0.2106	0.8241	0.6599	0.9805	0.8241	0.1759	0.8241	0.0195	0.9805
[76]	Used	84.40	0.1412	0.8440	0.7093	0.9827	0.8440	0.1560	0.8440	0.0173	0.9827

as well as on the contours of pigmented lesions. Table 5 presents the results of calculations of various methods for the quantitative evaluation of neural network systems. As a result of the modeling, it was found that the use of a modified cross-entropy loss function with the help of weight coefficients can improve the accuracy of neural network recognition and reduce the value of the loss function. All codes used for practical modeling are presented at the link [78]. The highest recognition accuracy of dermatological data was 85.20% and was obtained when testing the proposed multimodal neural network system that is sensitive to unbalanced data based on the DenseNet_161 architecture. When testing each of the proposed multimodal neural network architectures that are sensitive to unbalanced dermatological data, the recognition accuracy was higher than when testing the original multimodal neural network architectures. The increase in accuracy due to the multimodal approach using weights for unbalanced data ranged from 1.99 to 4.28 percentage points, depending on the selected pre-trained CNN. The smallest loss function index was 0.1294 and was obtained when testing a multimodal neural network system that is sensitive to unbalanced data based on the DenseNet_161 architecture. The value of the loss function of the proposed multimodal neural network systems with a modified cross-entropy loss function was in all cases lower than that of the original multimodal neural network architectures. The decrease in the loss function exponent was 0.0694-0.1670 depending on the selected pre-trained CNN.

For the statistical evaluation of the trained models, such performance measures as Specificity, Sensitivity, F-1 score, MCC, FNR, FPR, NPV and PPV. When evaluating intelligent systems for assisted dermatological diagnostics, sensitivity indicates how well the system can identify malignant skin lesions in patients who do have pigmentary oncopathology. The higher the sensitivity, the more reliable the intelligent medical system. When testing the proposed multimodal neural network systems for dermatological data recognition, it was found that the highest sensitivity index belongs to the proposed system based on the DenseNet_161 architecture with a modified cross-entropy loss function and is 0.8520. Specificity indicates how well the neural network system identifies patients with benign pigmented neoplasms. The best sensitivity index was obtained for a multimodal neural network system sensitive to unbalanced data based on the DenseNet_161 architecture and amounted to 0.9836. F-1 score is a measure of the evaluation of neural network systems and represents the harmonic mean of positive predictive value and sensitivity. The best F-1 score was obtained when testing the proposed neural network system with a modified loss function based on the DenseNet_161 architecture and amounted to 0.8520. At the same time, the statistical metric F-1 score is dependent on the ratio of positive and negative cases and cannot always correctly evaluate systems in which there is a clear imbalance of data. MCC is a more reliable measure of the statistical evaluation of systems with unbalanced data. A high MCC score indicates that the neural network system performs well in all four categories of the confusion matrix in proportion to the number of benign and malignant cases in the data set [79]. The best MCC score was 0.7168 and was obtained when evaluating a multimodal neural network system based on the DenseNet_161 architecture, which is sensitive to unbalanced data. FNR and FPR are the probability of false and true rejection of the null hypothesis as a result of testing a neural network system. PPV and NPV indicate the proportion of benign and malignant system test results that are truly benign and truly malignant. As a result of testing all trained neural network systems, the best result for all four indicators FNR, FPR, NPV and PPV was obtained from a neural network system based on the DenseNet_161 architecture, which is sensitive to unbalanced data and amounted to 0.1480, 0.0164, 0.9836 and 0.8520, respectively. For all the considered testing metrics, the systems trained using the modified cross-entropy loss function had a higher result than the original multimodal systems for recognizing pigmented skin lesions. The use of a modified cross-entropy loss function when training multimodal neural network systems made it possible to obtain classifiers that are sensitive to unbalanced dermatological data. Figures 11-13 show confusion matrices for testing multimodal neural network systems. Diagnostic categories are arranged in order of increasing risk and severity of the course of the disease.

As a result of the analysis of the confusion matrices in Figures 11-13, it can be concluded that the use of the modified cross-entropy loss function when training various multimodal neural network systems makes it possible to increase the accuracy of the classification of pigmented lesions in 10 diagnostically significant categories. When using the modified cross-entropy loss function during training, in all cases there was an increase in the number of recognized cases of melanoma. When using the DenseNet_161 neural network architecture, the number of correct melanoma classifications increased from 813 to 872 cases. When using

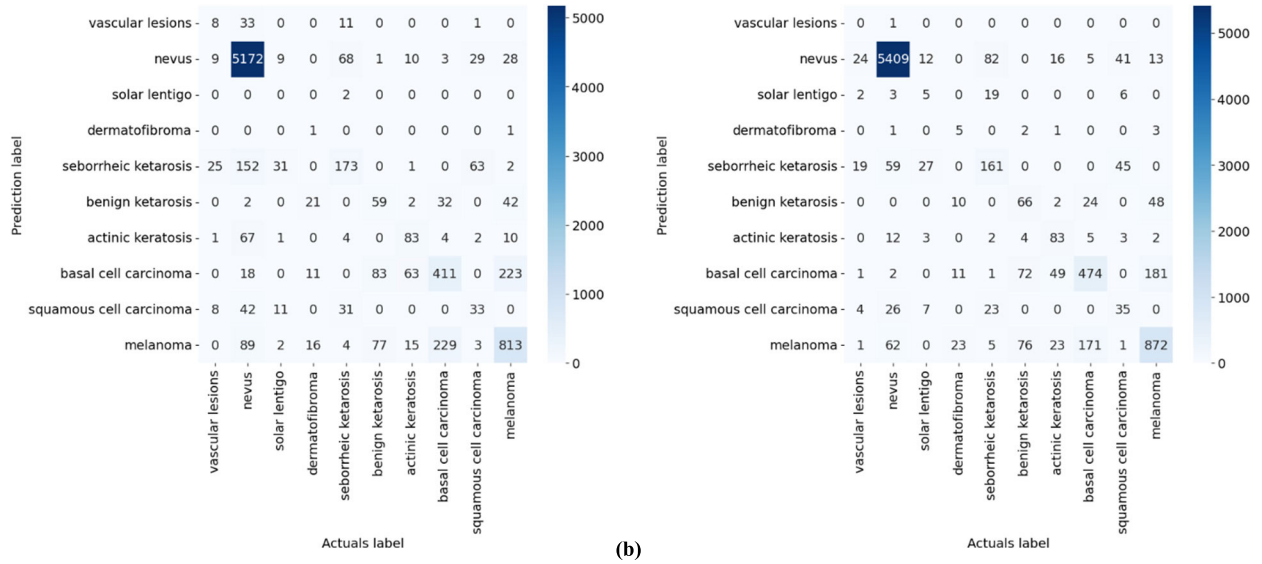

FIGURE 11. Confusion matrices as a result of testing a multimodal neural network system based on the DenseNet_161 architecture: a) original multimodal neural network system; b) multimodal neural network system with a modified cross-entropy loss function.

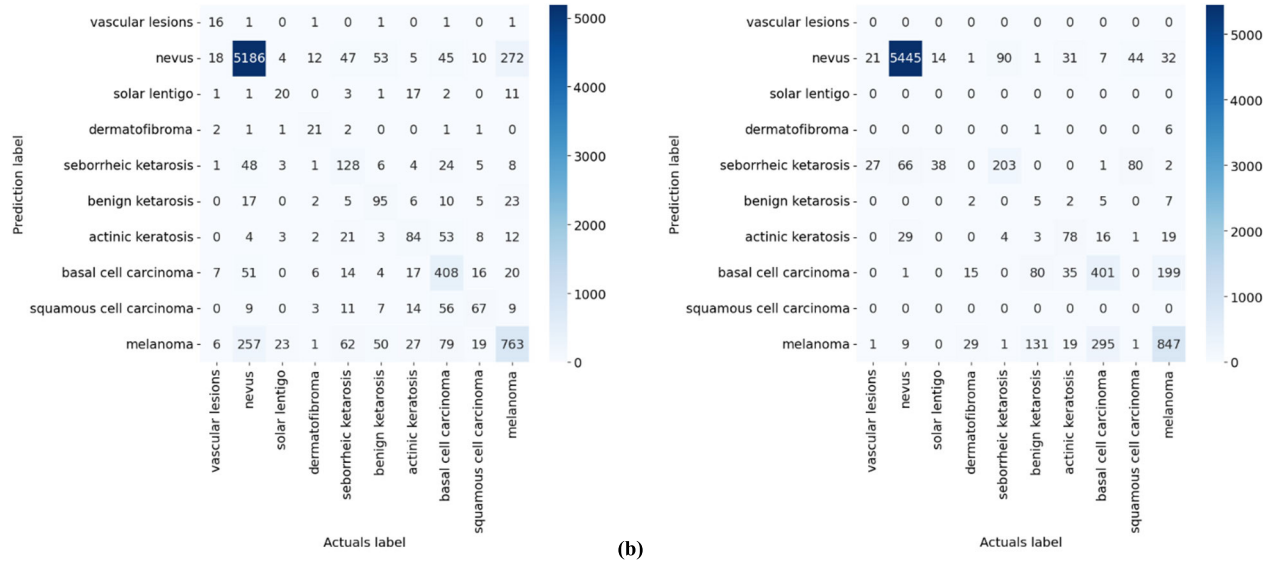

FIGURE 12. Confusion matrices as a result of testing a multimodal neural network system based on the Inception_v4 architecture: a) original multimodal neural network system; b) multimodal neural network system with a modified cross-entropy loss function.

TABLE 6. The results of testing multimodal neural network systems by the proposed methods of quantitative assessment for two categories.

CNN architecture	Loss function weights	Accuracy	Sp_{onco}	Se_{onco}	$F1_{onco}$	MCC_{onco}	FNR_{onco}	FPR_{onco}	PPV_{onco}	NPV_{onco}
DenseNet_161 [74]	Not used	0.9186	0.8982	0.9255	0.9445	0.7947	0.1018	0.0745	0.8025	0.9643
	Used	0.9354	0.9030	0.9463	0.9564	0.8328	0.0970	0.0537	0.8500	0.9666
Inception_v4 [75]	Not used	0.8808	0.7855	0.9128	0.9197	0.6886	0.2145	0.0872	0.7523	0.9128
	Used	0.9377	0.8980	0.9513	0.9579	0.8385	0.1020	0.0487	0.8631	0.9646
ResNeXt_50 [76]	Not used	0.9243	0.8755	0.9409	0.9488	0.8038	0.1245	0.0591	0.8345	0.9569
	Used	0.9385	0.8957	0.9532	0.9585	0.8401	0.1043	0.0468	0.8675	0.9639

the Inception_v4 neural network architecture, the number of correct melanoma recognitions increased from 763 to 847. When using the ResNeXt_50 architecture, the number of correct melanoma classifications increased from 874 to 961.

The predicted values obtained from the testing of each multiclass neural network system from the confusion matrices were grouped into two categories "benign" and "malignant" to calculate the proposed statistical estimates $F1_{onco}$, MCC_{onco}, Sp_{onco}, Se_{onco}, FNR_{onco}, FPR_{onco}, NPV_{onco} and PPV_{onco}. Table 6 presents the results of calculating the proposed statistical estimates when testing multimodal neural network systems. The highest accuracy rate for two groups

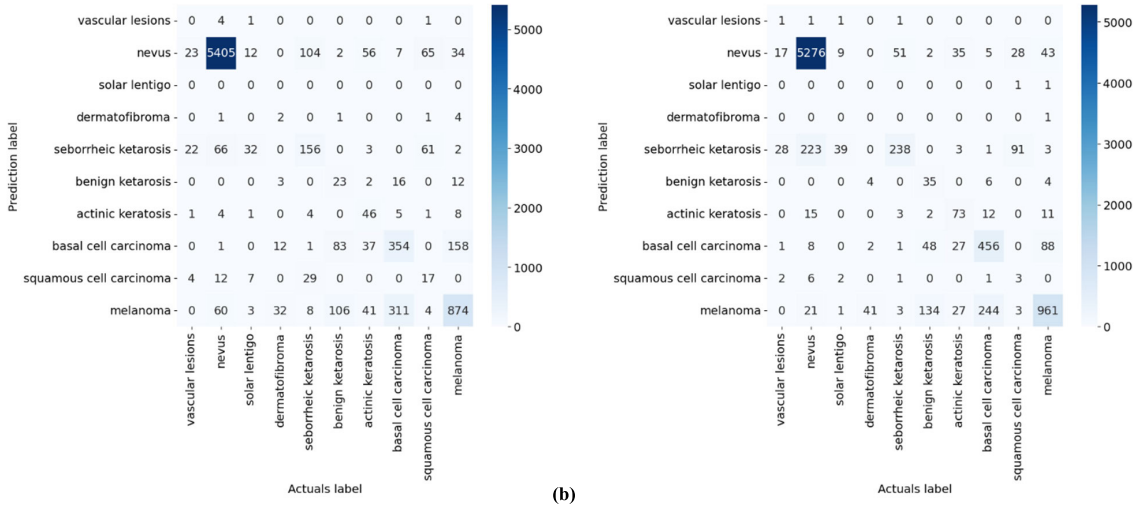

FIGURE 13. Confusion matrices as a result of testing a multimodal neural network system based on the ResNeXt_50 architecture: a) original multimodal neural network system; b) multimodal neural network system with a modified cross-entropy loss function.

of ''benign'' and ''malignant'' pigmented skin lesions was obtained with a multiclass neural network system based on the ResNeXt_50 architecture, trained using weight coefficients and amounted to 93.85%. The best indicator of the FNR_{onco} was obtained for a multiclass neural network system based on the DenseNet_161 architecture, trained using weight coefficients and amounted to 0.0970. In all cases, the indicator FNR_{onco} was lower for multimodal neural network systems trained using weight coefficients. The use of weighting factors for unbalanced data significantly reduced the occurrence of false negative predictions when a pigmented lesion from the malignant categories was recognized as a pigmented lesion from the benign categories. Figures 14-15 show the confusion matrices for testing multimodal neural network systems in two categories.

As a result of the analysis of confusion matrices in Figures 14-15, it can be concluded that the use of the modified cross-entropy loss function when training various multimodal neural network systems can reduce the number of false positive and false negative predictions. For intelligent systems of medical auxiliary diagnostics, reducing the percentage of false negative predictions is a critical task. The greatest result in the reduction of cases of false negative prediction was obtained when comparing multimodal neural network systems based on the Inception_v4 architecture and amounted to 234 cases. The use of the modified cross-entropy loss function reduced the number of cases of false-negative recognition of pigmented skin lesions by 10 cases for the architecture based on DenseNet_161, by 234 cases for the architecture based on Inception_v4 and by 42 cases for the architecture based on ResNeXt_50.

As a result of calculations for the McNemar test in Figure 16, it was found that the use of the modified cross-entropy loss function at the training stage of the neural network system made it possible to increase the number of correct recognitions. In 892-1149 cases, the multimodal system, sensitive to unbalanced data, correctly classified skin pigments. When classifying these same cases, the original multimodal neural network system made mistakes. In 200-217 cases, the recognition results of a multimodal system sensitive to unbalanced data were incorrect compared to the original neural network system.

Due to the more severe punishment when training a multimodal neural network, it was possible to obtain a neural network system that is sensitive to unbalanced data. However, the proposed system cannot be used as an independent diagnostic tool due to the risk of false negative loss. However, the use of weighting factors for unbalanced data has significantly reduced the occurrence of false negative predictions in neural network systems based on various convolutional architectures.

IV. DISCUSSION

The paper presents a multimodal neural network system with a modified cross-entropy loss function, sensitive to unbalanced heterogeneous dermatological data. The accuracy of the proposed neural network system based on the DenseNet_161 convolutional architecture was 85.20%. The system analyzes heterogeneous dermatological data represented by images of pigmented skin lesions and such metadata as gender, age and location of pigmented lesions on the body. At the same time, the educational dermatological data available in the public domain are highly unbalanced towards ''benign'' categories. The modification of the cross-entropy loss function made it possible to overcome the data imbalance and achieve higher accuracy compared to the results of testing the original multimodal systems, as well as compared to the results of similar systems for detecting malignant skin lesions. Table 7 compares the accuracy values for recognition of pigmented skin lesions when testing the proposed system, which is sensitive to unbalanced data, with the accuracy values for testing similar multimodal systems. The table shows the test accuracy of the considered systems,

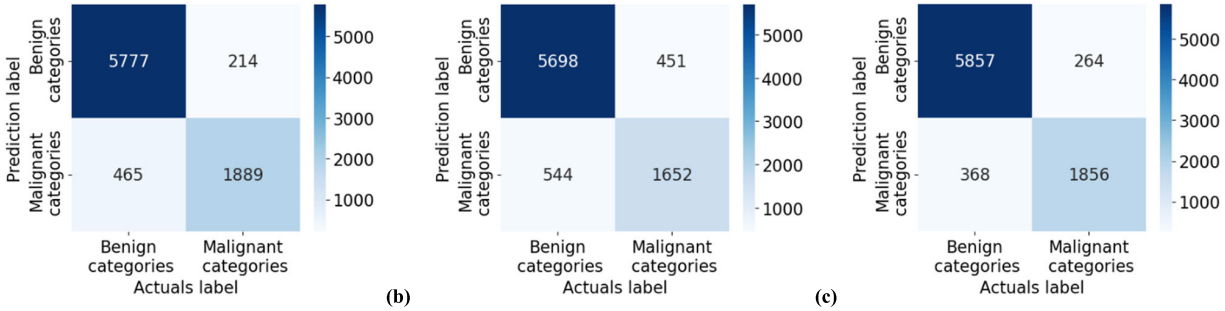

FIGURE 14. Confusion matrices in two categories as a result of testing the original multimodal neural network system based on architectures: a) DenseNet_161; b) Inception_v4; c) ResNeXt_50.

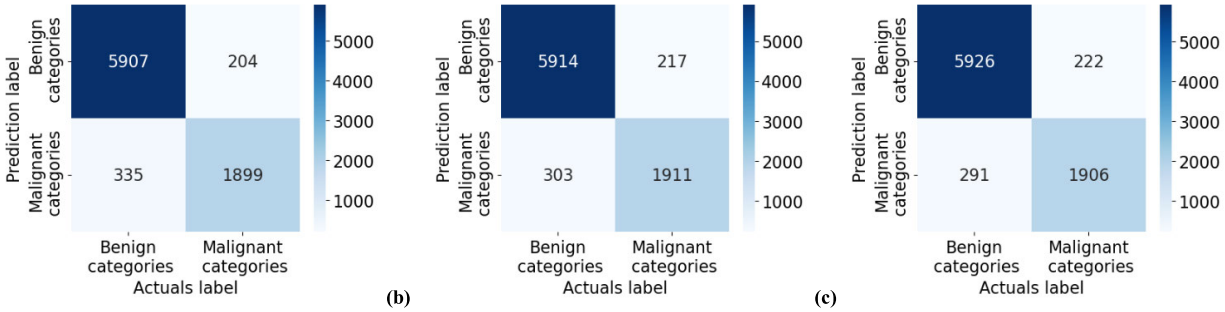

FIGURE 15. Confusion matrices in two categories as a result of testing a multimodal neural network system modified with a cross-entropy loss function based on architectures: a) DenseNet_161; b) Inception_v4; c) ResNeXt_50.

(a)

	Multimodal neural network system with a modified cross-entropy loss function, **Correct**	Multimodal neural network system with a modified cross-entropy loss function, **Incorrect**	Total:
Original multimodal neural network system, **Correct**	6531	217	6748
Original multimodal neural network system, **Incorrect**	1149	827	1976
Total:	7680	1044	

X^2 McNemar	652.420
Statistical Significance:	0.001

(b)

	Multimodal neural network system with a modified cross-entropy loss function, **Correct**	Multimodal neural network system with a modified cross-entropy loss function, **Incorrect**	Total:
Original multimodal neural network system, **Correct**	6759	211	6970
Original multimodal neural network system, **Incorrect**	894	803	1697
Total:	7653	1014	

X^2 McNemar	360.749
Statistical Significance:	0.001

(c)

	Multimodal neural network system with a modified cross-entropy loss function, **Correct**	Multimodal neural network system with a modified cross-entropy loss function, **Incorrect**	Total:
Original multimodal neural network system, **Correct**	6973	200	7173
Original multimodal neural network system, **Incorrect**	892	904	1796
Total:	7865	1104	

X^2 McNemar	394.804
Statistical Significance:	0.001

FIGURE 16. Classification tables for testing multimodal neural network systems for recognizing pigmented skin lesions for McNemar analysis based on architectures: a) DenseNet_161; b) Inception_v4; c) ResNeXt_50.

indicated in the research papers. All considered models from the presented studies cannot be compared directly, since training and further testing were carried out on different sets of dermatological data.

The work [80] presents a method for intelligent recognition of heterogeneous data, such as clinical images and metadata. Modeling was performed on a database of 2917 clinical cases from five classes (1127 examples of nevi, 727 examples of melanoma, 647 examples of basal cell carcinoma, 273 examples of squamous cell carcinoma, 143 examples of pigmented benign keratoses). As a result, the average test accuracy for multi-class classification of the ResNet-50 multimodal neural network architecture was 71.90%, which is 12.50 percentage points lower than the accuracy results of the proposed multimodal system with a similar ResNeXt_50 architecture trained with a modified cross-entropy loss function. This result is 13.30 percentage points lower than the test accuracy of the proposed multimodal system with the best DenseNet_161 architecture in terms of accuracy. The use of more training data, as well as the use of preprocessing methods and modification of the cross-entropy loss function, made it possible to significantly increase the accuracy of recognition of dermatological data compared to similar systems.

The work [81] presents a multimodal neural network system CAFNet, which analyzes such heterogeneous dermatological data as dermoscopic and clinical images. Modeling of the considered system was carried out by the seven-point checklist dataset. This set contained 1011 examples of pigmented neoplasms divided into five classes (575 examples of nevi, 252 examples of melanoma, 97 examples of mixed malignant category (MISC), 42 examples of basal cell carcinoma, 45 examples of seborrheic keratosis). Each data sample included a pair of clinical and dermoscopic images that

TABLE 7. Accuracy values in testing various multimodal neural network systems for recognizing pigmented skin lesions.

Multimodal neural network system for recognizing pigmented skin lesions		Accuracy of recognition of pigmented neoplasms of the skin, %
Known neural network systems	[80]	71.90
	[81]	76.80
	[82]	78.80
	[83]	80.39
	[84]	80.42
	[85]	83.93
The proposed multimodal neural network system based on the DenseNet_161 architecture		85.20

reflected the same pigmented skin lesion in a patient, seven-point criteria labels, and a seven-point score. The CAFNet system uses two architectures for feature extraction from dermoscopic and clinical images and a neural network architecture for feature analysis. The results show that CAFNet achieves an average accuracy of 76.80% on a test dataset of 7 diagnostically relevant categories of pigmented skin lesions. The accuracy of the CAFNet neural network system is 7.39% compared to the accuracy of recognizing pigmented skin lesions using the ResNet-50 CNN. Despite a significant increase in the recognition accuracy of pigmented skin lesions when using heterogeneous visual data, the CAFNet test results are 8.40 percentage points lower than those of the proposed multimodal neural network system with a modified cross-entropy loss function based on the DenseNet_161 architecture. The joint use of visual data and metadata of patients made it possible to identify additional relationships between the diagnosis and pigmented neoplasm, thereby increasing the accuracy of intelligent diagnostics. At the same time, the use of the input data pre-processing stage also significantly improved the quality of the information processed by the artificial intelligence system.

The study [82] analyzed the impact of patient clinical information on skin cancer detection using deep learning models. The authors of the work presented a approach to combining image and clinical features using various CNN models. First, image features were extracted using CNN. These features were then reduced using a reduction block and further merged with clinical metadata. The combined signal was fed to the classification layer, which outputs the final classifier response. The simulation was carried out on a data set consisting of 1612 cases divided into 6 categories. The best result in terms of recognition accuracy in the work under consideration was obtained using an artificial intelligence system based on the ResNet-50 architecture and amounted to 78.80%, which is 6.40 percentage points lower than that of the proposed multimodal system based on the DenseNet_161 architecture, sensitive to unbalanced data. At the same time, one of the significant drawbacks of the system based on the ResNet-50 architecture from [82] is the small number of examples in the data set for modeling, which can have a significant impact on the reliability of the results obtained and the further stability of the system.

The authors in [83] proposed an automatic method for classifying pigmented skin lesions using a CNN deep learning model. To improve the classification performance of the CNN, both images and patient statistics were used in the modeling process. Simulation results performed on the publicly available HAM10000 dataset showed that a standard CNN model based on the AlexNet architecture could classify skin lesions with an accuracy of 79.29%, while the authors' proposed multimodal approach improved the accuracy to 80.39%. The resulting test accuracy of the multimodal model under consideration is 4.81 percentage points lower than the accuracy of the developed multimodal intelligent system based on the DenseNet_161 architecture, which is sensitive to unbalanced data. A significant drawback of the multimodal model under consideration is the absence of a pre-processing stage for visual data, which makes it possible to prepare images for further feature extraction and can significantly increase the accuracy of the intelligent system.

The work [84] presents a multi-mode data fusion diagnostic network MDFNet, which combines heterogeneous features of clinical skin images and clinical data of patients. Modeling of the system under consideration was carried out on the PAD-UFES-20 data set, consisting of 2298 examples of six different types of pigmented skin lesions. Each example was diagnosed jointly by several experienced dermatologists. The actinic keratosis (ACK) class included 730 examples, the basal cell carcinoma of skin (BCC) class included 845 examples, the malignant melanoma (MEL) class included 52 examples, the melanocytic nevus of skin (NEV) class consisted of 244 examples, the squamous cell carcinoma (SCC) class included 192 examples and the seborrheic keratosis (SEK) class included 235 examples. In addition to clinical images, the data set also included 21 clinical data items such as patient age, gender, cancer history, and family history. The experimental results showed that the MDFNet system has an accuracy of 80.42% on the test data, which is about 9% higher than the accuracy of the neural network model using only dermatological images. Modeling of the system was carried out on 2298 clinical cases, divided into six categories of pigmented skin lesions. The authors of the work used ResNet_50 and DenseNet_121 as neural network architectures. Test evaluation of MDFNet based on the ResNet_50 architecture made it possible to obtain a classification accuracy of 77.11%, which is 7.29 percentage points lower than the accuracy of the proposed multimodal system based on a CNN ResNeXt_50. Test evaluation of MDFNet based on the DenseNet_121 architecture showed an accuracy of 80.42%, which is 4.78 percentage points lower than the accuracy of the proposed multimodal system based on the similar CNN DenseNet_161. Training using a modified cross-entropy loss function using weighting coefficients made it possible to obtain a classifier that is sensitive to unbalanced dermatological data and to reduce the frequency of false negative loss, in which malignant pigmented neoplasms are recognized as benign.

The work [85] presented a mobile intelligent system for recognizing pigmented skin tumors based on the MobileNetV2 architecture. In the study under review, simulations were carried out using the HAM10000 dataset. The classification of the mobile intelligent system was carried out into two categories: ''benign'' and ''malignant'' pigmented skin lesions. No preprocessing was applied to the data except for balancing, resizing, and data augmentation through affine transformations. The best performance was achieved using a simplified MobileNetV2 combined with an attention coordination mechanism with a total of less than a million parameters, and the accuracy was 83.93%. The considered mobile intelligent system recognizes pigmented skin tumors 1.27 percentage points less accurately compared to the proposed multimodal intelligent system based on the DenseNet_161 architecture, trained using a modified cross-entropy loss function. In addition, a mobile intelligent system requires a mobile device with this architecture located on the device itself, while the developed multimodal intelligent system is capable of working as a web application hosted on a server and only requires Internet access from any digital device with the ability to download images.

The work [86] presented an ensemble system that won the 2020 ISIC Melanoma Classification Challenge. The system makes predictions in two categories: ''benign'' and ''malignant'' pigmented skin tumors. In this case, the class of ''benign'' pigmented neoplasms is further divided into 8 subclasses. Part of the neural network architectures of the considered ensemble model includes multimodal analysis of heterogeneous dermatological data, such as visual data and metadata. The authors found that models that used metadata in conjunction with visual data provided higher accuracy and stability. The creation of an ensemble system that includes the most highly accurate CNN architectures is one of the promising areas for further research.

A fast intelligent recognition approach using EfficientNet for diagnosing twenty types of skin diseases was presented in [87]. This system recognizes skin diseases such as lupus, psoriasis, herpes, lichen, dermatitis and others. Among the 8 EfficientNet models from B0 to B7, the EfficientNet-B7 based model achieved the highest accuracy of 97.10% on the test dataset. Such intelligent systems can become a promising direction for the development of ensemble models, in which skin pathology can be recognized as a ''pigmented neoplasm'' or ''skin disease'' with further more accurate classification into various categories.

The proposed multimodal system trained with a modified cross-entropy loss function significantly exceeds the accuracy of visual analysis methods used by oncol dermatologists. A comparison of the accuracy of classification of pigmented skin lesions in dermatologists with different levels of experience and an artificial intelligence system was presented in [88], [89], and [90] by a computer program using an artificial algorithm. skin neoplasms. The proposed system of multimodal neural network analysis of heterogeneous dermatological data can significantly reduce the risk of false negative loss, when a potentially malignant pigmented formation can be recognized as benign. This system takes into account such statistical factors as gender, age, localization of pigmented formation. However, the developed multimodal neural network system, which is sensitive to unbalanced data, cannot replace the decisive opinion of a specialist. The developed system cannot be used as an independent diagnostic tool due to the possible risk of false negative loss. It is assumed that dermatologists-oncologists will be able to use the proposed system only as a high-precision auxiliary diagnostic tool to help make a medical decision. Before the introduction of this system into the clinical setting, a number of tests are required on real dermatological cases under the supervision of dermatologists.

Promising direction for further research is the construction of more complex ensemble systems for neural network analysis of dermatological data. Another promising area for further research is the introduction of segmentation at the stage of pre-processing of visual data. Semantic segmentation will make it possible to highlight the contour of a pigmented neoplasm, the distortion of which is a diagnostic morphological manifestation of oncopathology. The development of web applications and computer programs for implementation in the healthcare sector as auxiliary tools for diagnosing oncopathologies is also relevant.

V. CONCLUSION

As a result of the study, an intelligent classification system for pigmented skin lesions was developed based on a multimodal neural network that is sensitive to unbalanced data. The accuracy of the proposed system was 85.20% for the architecture based on the DenseNet_161 CNN. In all cases, the recognition accuracy of dermatological data in systems sensitive to unbalanced data was higher compared to the original multimodal neural network systems. The increase in accuracy due to the multimodal approach using weights for unbalanced data ranged from 1.99 to 4.28 percentage points, depending on the selected pre-trained CNN. The value of the error function for the developed system based on the DenseNet_161 architecture was 0.1294. In all cases, there was a decrease in the value of the error function in systems sensitive to unbalanced data compared to the original neural network architectures. The decrease in the error function exponent was 0.0694-0.1670, depending on the selected pre-trained CNN. The use of the visual data pre-processing stage, as well as the modified cross-entropy loss function during training, made it possible to increase the accuracy of the multimodal intelligent system and increase the sensitivity to unbalanced dermatological data. At the same time, a significant reduction in the number of false negative and false negative predictions was found, which is a critical task for recognition systems of malignant skin lesions. Each of the presented elements of improving the accuracy of classification, such as cleaning hair structures, analyzing heterogeneous data, the balanced learning method by modifying the cross-entropy loss function, enters into synergy with other methods introduced into

the system. Due to the emerging synergy, there was a significant increase in forecasting accuracy and a decrease in the number of false negative responses. Since the modeling was carried out on independently selected data from the open ISIC archive, the developed architectures were trained and tested on various types of dermatological images. The proposed multimodal systems, sensitive to unbalanced data, are more resistant to varying image quality, color, resolution, and the presence of different noise structures.

The main limitation of using the proposed multimodal neural network system with a modified cross-entropy loss function sensitive to unbalanced heterogeneous dermatological data is that specialists can only use the system as an additional diagnostic tool. The developed system cannot be used as an independent diagnostic tool due to the possible risk of false negative loss. However, the proposed system can be used as a highly accurate auxiliary tool to assist in making a medical decision. The introduction of such high-precision systems for automated analysis of pigmented skin lesions will reduce the consumption of financial and labor resources involved in the medical industry, as well as increase the chance of early detection of pigmented oncopathologies.

When introducing artificial intelligence systems as auxiliary medical tools, cases may arise of diagnosing rare pigmented skin tumors that the intelligent system has not been trained to recognize, or the appearance of rare localizations of pigmented lesions, for example, on the palms. All this can lead to a false diagnosis and a decrease in the performance of the neural network model [91]. Thus, another limitation when using the developed multimodal intelligent system for analyzing unbalanced data is the insufficient level of generalizability of the model for rare cases of pigmented skin tumors on which the model was not trained.

Another possible limitation in the application of the proposed multimodal intelligent system is the presence of several pigmented skin tumors in the visual data. In these cases, classification will be made only for the most contrasting and largest pigmented lesion due to the ability of CNN to highlight the most significant features. A possible solution to this problem is the introduction of detection methods with further segmentation to identify each pigmented neoplasm as a separate sample for classification. The lack of detection methods also entails the problem of classifying images in which there are no pigmented skin tumors. Since the training of the proposed multimodal intelligent systems was carried out only in 10 diagnostically significant categories, an image without pigmented skin lesions will in any case be classified as one of the possible skin neoplasms. Such false diagnoses require careful verification of recognized images and classification results by doctors or specialists.

APPENDIX A

Appendix A shows the training and testing graphs of the proposed multimodal neural network systems based on various CNN architectures.

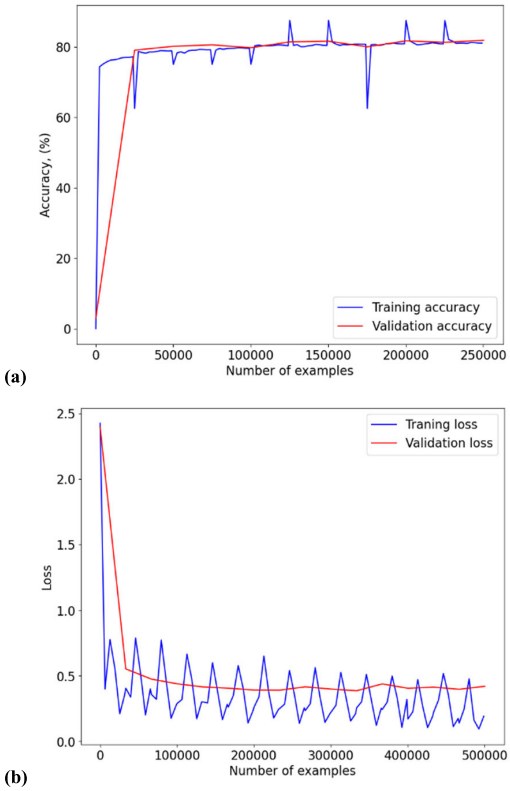

FIGURE 17. Graph of training and validation results of the multimodal neural network system for analyzing heterogeneous dermatological data CNN DenseNet_161 without using weight coefficients for the loss function: a) recognition accuracy; b) loss function.

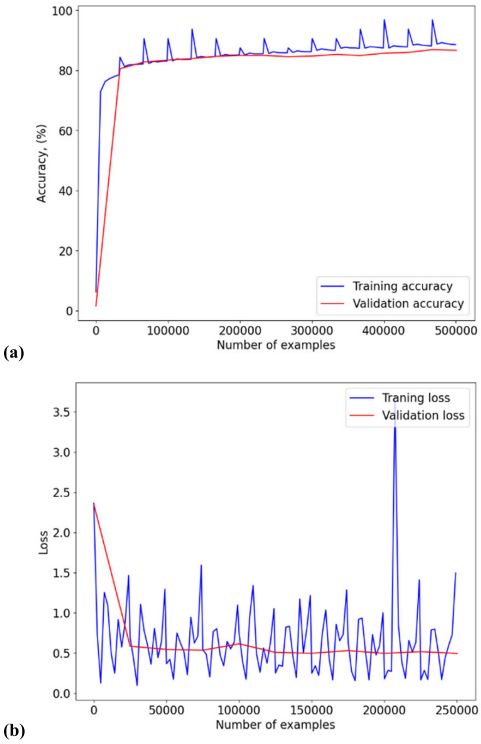

FIGURE 18. Graph of training and validation results of the multimodal neural network system for analyzing heterogeneous dermatological data CNN DenseNet_161 using weight coefficients for the loss function: a) recognition accuracy; b) loss function.

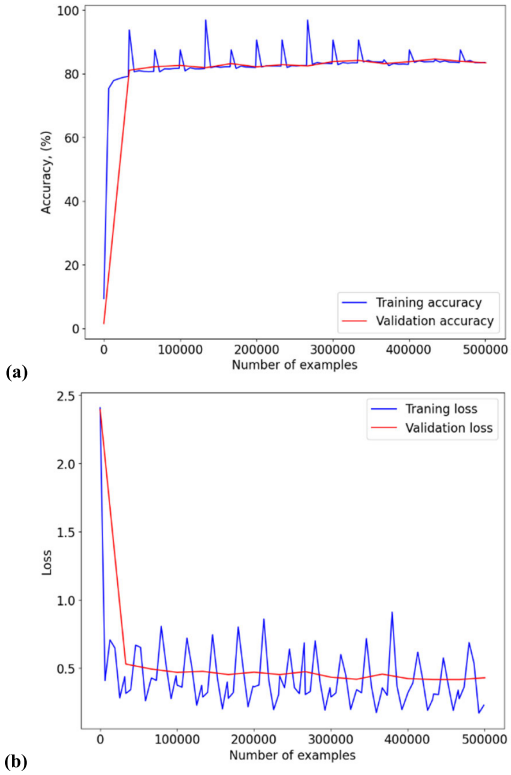

FIGURE 19. Graph of training and validation results of the multimodal neural network system for analyzing heterogeneous dermatological data CNN Inception_v4 without using weight coefficients for the loss function: a) recognition accuracy; b) loss function.

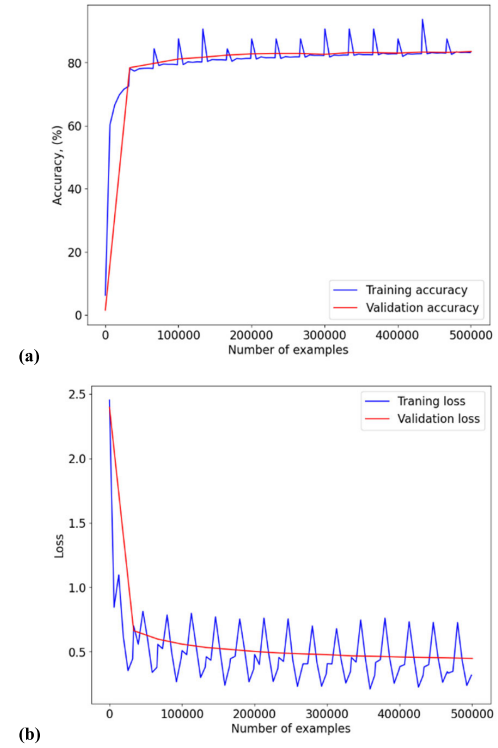

FIGURE 20. Graph of training and validation results of the multimodal neural network system for analyzing heterogeneous dermatological data CNN Inception_v4 using weight coefficients for the loss function: a) recognition accuracy; b) loss function.

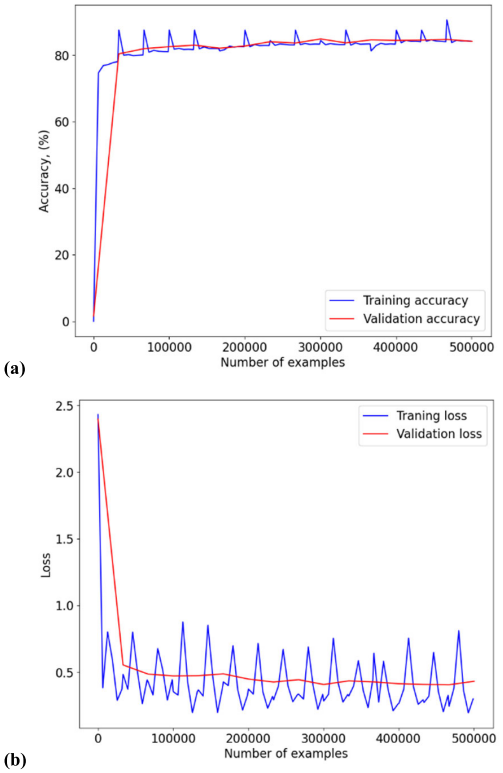

FIGURE 21. Graph of training and validation results of the multimodal neural network system for analyzing heterogeneous dermatological data CNN ResNeXt_50 without using weight coefficients for the loss function: a) recognition accuracy; b) loss function.

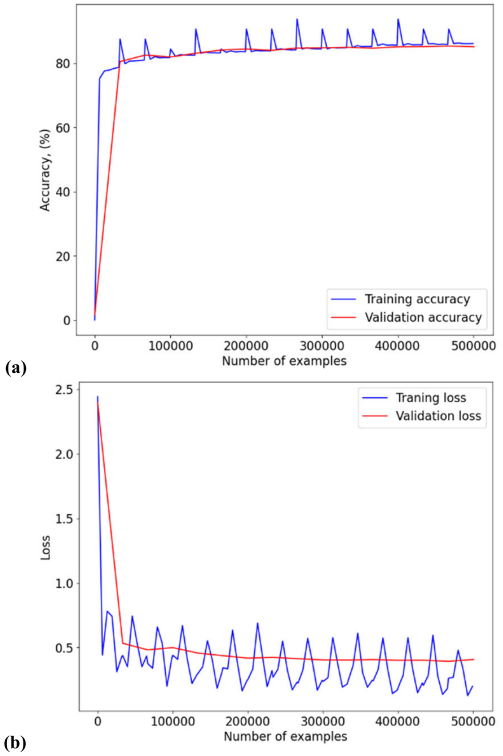

FIGURE 22. Graph of training and validation results of the multimodal neural network system for analyzing heterogeneous dermatological data CNN ResNeXt_50 using weight coefficients for the loss function: a) recognition accuracy; b) loss function.

ACKNOWLEDGMENT

The authors express their gratitude to North-Caucasus Federal University for supporting the competition of projects of research teams and individual scientists of North-Caucasus Federal University.

REFERENCES

[1] Z. Apalla, A. Lallas, E. Sotiriou, E. Lazaridou, and D. Ioannides, "Epidemiological trends in skin cancer," *Dermatol. Practical Conceptual*, vol. 7, no. 2, p. 1, Apr. 2017, doi: 10.5826/dpc.0702a01.

[2] W. Hu, L. Fang, R. Ni, H. Zhang, and G. Pan, "Changing trends in the disease burden of non-melanoma skin cancer globally from 1990 to 2019 and its predicted level in 25 years," *BMC Cancer*, vol. 22, no. 1, pp. 1–11, Dec. 2022, doi: 10.1186/s12885-022-09940-3.

[3] K. Saginala, A. Barsouk, J. S. Aluru, P. Rawla, and A. Barsouk, "Epidemiology of melanoma," *Med. Sci.*, vol. 9, no. 4, p. 63, Oct. 2021, doi: 10.3390/medsci9040063.

[4] N. R. Kurtansky, S. W. Dusza, A. C. Halpern, R. I. Hartman, A. C. Geller, A. A. Marghoob, V. M. Rotemberg, and M. A. Marchetti, "An epidemiologic analysis of melanoma overdiagnosis in the United States, 1975–2017," *J. Investigative Dermatol.*, vol. 142, no. 7, pp. 1804–1811, Jul. 2022, doi: 10.1016/j.jid.2021.12.003.

[5] M. Ciąyńska et al., "The incidence and clinical analysis of non-melanoma skin cancer," *Sci. Rep.*, vol. 11, no. 1, pp. 1–10, Feb. 2021, doi: 10.1038/s41598-021-83502-8.

[6] *Melanoma Awareness Month 2022—IARC*. Accessed: Dec. 14, 2022. [Online]. Available: https://www.iarc.who.int/news-events/melanoma-awareness-month-2022/

[7] M. Arnold, D. Singh, M. Laversanne, J. Vignat, S. Vaccarella, F. Meheus, A. E. Cust, E. de Vries, D. C. Whiteman, and F. Bray, "Global burden of cutaneous melanoma in 2020 and projections to 2040," *JAMA Dermatol.*, vol. 158, no. 5, p. 495, May 2022, doi: 10.1001/jamadermatol.2022.0160.

[8] B. S. Allais, M. Beatson, H. Wang, S. Shahbazi, L. Bijelic, S. Jang, and S. Venna, "Five-year survival in patients with nodular and superficial spreading melanomas in the U.S. population," *J. Amer. Acad. Dermatol.*, vol. 84, no. 4, pp. 1015–1022, Apr. 2021, doi: 10.1016/j.jaad.2020.11.047.

[9] D. Schadendorf, A. C. J. van Akkooi, C. Berking, K. G. Griewank, R. Gutzmer, A. Hauschild, A. Stang, A. Roesch, and S. Ugurel, "Melanoma," *Lancet*, vol. 392, no. 10151, pp. 971–984, Sep. 2018, doi: 10.1016/s0140-6736(18)31559-9.

[10] A. Lideikaitė, J. Mozūraitienė, and S. Letautienė, "Analysis of prognostic factors for melanoma patients," *Acta medica Lituanica*, vol. 24, no. 1, pp. 25–34, Apr. 2017, doi: 10.6001/actamedica.v24i1.3460.

[11] T. Amaral and C. Garbe, "Non-melanoma skin cancer: New and future synthetic drug treatments," *Expert Opinion Pharmacotherapy*, vol. 18, no. 7, pp. 689–699, May 2017, doi: 10.1080/14656566.2017.1316372.

[12] I. M. Leigh, "Progress in skin cancer: The U.K. experience," *Brit. J. Dermatol.*, vol. 171, no. 3, pp. 443–445, Sep. 2014, doi: 10.1111/bjd.13258.

[13] M. A. Linares, A. Zakaria, and P. Nizran, "Skin cancer," *Primary Care, Clinics Office Pract.*, vol. 42, no. 4, pp. 645–659, Dec. 2015, doi: 10.1016/j.pop.2015.07.006.

[14] A. M. Juszczak, U. Wöelfle, and M. Tomczyk, "Skin cancer, including related pathways and therapy and the role of luteolin derivatives as potential therapeutics," *Medicinal Res. Rev.*, vol. 42, no. 4, pp. 1423–1462, Jul. 2022, doi: 10.1002/med.21880.

[15] V. Collier, M. Musicante, T. Patel, and F. Liu-Smith, "Sex disparity in skin carcinogenesis and potential influence of sex hormones," *Skin Health Disease*, vol. 1, no. 2, p. e27, Jun. 2021, doi: 10.1002/ski2.27.

[16] Z. Apalla, P. Calzavara-Pinton, A. Lallas, G. Argenziano, A. Kyrgidis, S. Crotti, F. Facchetti, P. Monari, and G. Gualdi, "Histopathological study of perilesional skin in patients diagnosed with nonmelanoma skin cancer," *Clin. Experim. Dermatol.*, vol. 41, no. 1, pp. 21–25, Jan. 2016, doi: 10.1111/ced.12713.

[17] C. Ring, N. Cox, and J. B. Lee, "Dermatoscopy," *Clinics Dermatol.*, vol. 39, no. 4, pp. 635–642, Jul. 2021, doi: 10.1016/j.clindermatol.2021.03.009.

[18] N. Nami, E. Giannini, M. Burroni, M. Fimiani, and P. Rubegni, "Teledermatology: State-of-the-art and future perspectives," *Expert Rev. Dermatol.*, vol. 7, no. 1, pp. 1–3, Feb. 2012, doi: 10.1586/edm.11.79.

[19] C. Sinz et al., "Accuracy of dermatoscopy for the diagnosis of nonpigmented cancers of the skin," *J. Amer. Acad. Dermatol.*, vol. 77, no. 6, pp. 1100–1109, Dec. 2017, doi: 10.1016/j.jaad.2017.07.022.

[20] M. E. Vestergaard, P. Macaskill, P. E. Holt, and S. W. Menzies, "Dermoscopy compared with naked eye examination for the diagnosis of primary melanoma: A meta-analysis of studies performed in a clinical setting," *Brit. J. Dermatol.*, vol. 159, no. 3, pp. 669–676, Jun. 2008, doi: 10.1111/j.1365-2133.2008.08713.x.

[21] H. A. Haenssle et al., "Man against machine: Diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists," *Ann. Oncol.*, vol. 29, no. 8, pp. 1836–1842, Aug. 2018, doi: 10.1093/annonc/mdy166.

[22] L. Brochez, E. Verhaeghe, E. Grosshans, E. Haneke, G. Piérard, D. Ruiter, and J. Naeyaert, "Inter-observer variation in the histopathological diagnosis of clinically suspicious pigmented skin lesions," *J. Pathol.*, vol. 196, no. 4, pp. 459–466, Apr. 2002, doi: 10.1002/path.1061.

[23] S. Lodha, S. Saggar, J. T. Celebi, and D. N. Silvers, "Discordance in the histopathologic diagnosis of difficult melanocytic neoplasms in the clinical setting," *J. Cutaneous Pathol.*, vol. 35, no. 4, pp. 349–352, Apr. 2008, doi: 10.1111/j.1600-0560.2007.00970.x.

[24] S. Haggenmüller et al., "Skin cancer classification via convolutional neural networks: Systematic review of studies involving human experts," *Eur. J. Cancer*, vol. 156, pp. 202–216, Oct. 2021, doi: 10.1016/j.ejca.2021.06.049.

[25] Y. Yan, S. Hong, W. Zhang, and H. Li, "Artificial intelligence in skin diseases: Fulfilling its potentials to meet the real needs in dermatology practice," *Health Data Sci.*, vol. 2022, pp. 1–2, Jan. 2022, doi: 10.34133/2022/9791467.

[26] J. Wiens, S. Saria, M. Sendak, M. Ghassemi, V. X. Liu, F. Doshi-Velez, K. Jung, K. Heller, D. Kale, M. Saeed, P. N. Ossorio, S. Thadaney-Israni, and A. Goldenberg, "Author correction: Do no harm: A roadmap for responsible machine learning for health care," *Nature Med.*, vol. 25, no. 10, p. 1627, Oct. 2019, doi: 10.1038/s41591-019-0609-x.

[27] W. Gao, L. Chen, and T. Shang, "Stream of unbalanced medical big data using convolutional neural network," *IEEE Access*, vol. 8, pp. 81310–81319, 2020, doi: 10.1109/ACCESS.2020.2991202.

[28] H. Yang, X. Li, H. Cao, Y. Cui, Y. Luo, J. Liu, and Y. Zhang, "Using machine learning methods to predict hepatic encephalopathy in cirrhotic patients with unbalanced data," *Comput. Methods Programs Biomed.*, vol. 211, Nov. 2021, Art. no. 106420, doi: 10.1016/j.cmpb.2021.106420.

[29] F. J. Moreno-Barea, J. M. Jerez, and L. Franco, "Improving classification accuracy using data augmentation on small data sets," *Expert Syst. Appl.*, vol. 161, Dec. 2020, Art. no. 113696, doi: 10.1016/j.eswa.2020.113696.

[30] M. Buda, A. Maki, and M. A. Mazurowski, "A systematic study of the class imbalance problem in convolutional neural networks," *Neural Netw.*, vol. 106, pp. 249–259, Oct. 2018, doi: 10.1016/j.neunet.2018.07.011.

[31] I. Czarnowski, "Weighted ensemble with one-class classification and oversampling and instance selection (WECOI): An approach for learning from imbalanced data streams," *J. Comput. Sci.*, vol. 61, May 2022, Art. no. 101614, doi: 10.1016/j.jocs.2022.101614.

[32] U. Bhowan, M. Johnston, M. Zhang, and X. Yao, "Evolving diverse ensembles using genetic programming for classification with unbalanced data," *IEEE Trans. Evol. Comput.*, vol. 17, no. 3, pp. 368–386, Jun. 2013, doi: 10.1109/TEVC.2012.2199119.

[33] A. Mikolajczyk and M. Grochowski, "Data augmentation for improving deep learning in image classification problem," in *Proc. Int. Interdiscipl. PhD Workshop (IIPhDW)*, May 2018, pp. 117–122, doi: 10.1109/IIPHDW.2018.8388338.

[34] N. Chen, Z. Xu, Z. Liu, Y. Chen, Y. Miao, Q. Li, Y. Hou, and L. Wang, "Data augmentation and intelligent recognition in pavement texture using a deep learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 12, pp. 25427–25436, Dec. 2022, doi: 10.1109/TITS.2022.3140586.

[35] P. Sadhukhan and S. Palit, "Reverse-nearest neighborhood based oversampling for imbalanced, multi-label datasets," *Pattern Recognit. Lett.*, vol. 125, pp. 813–820, Jul. 2019, doi: 10.1016/j.patrec.2019.08.009.

[36] M. Koziarski, "Radial-based undersampling for imbalanced data classification," *Pattern Recognit.*, vol. 102, Jun. 2020, Art. no. 107262, doi: 10.1016/j.patcog.2020.107262.

[37] Y. Zhu, C. Jia, F. Li, and J. Song, "Inspector: A lysine succinylation predictor based on edited nearest-neighbor undersampling and adaptive synthetic oversampling," *Anal. Biochem.*, vol. 593, Mar. 2020, Art. no. 113592, doi: 10.1016/j.ab.2020.113592.

[38] M. Kubus, "Evaluation of resampling methods in the class unbalance problem," *Econometrics*, vol. 24, no. 1, pp. 39–50, 2020, doi: 10.15611/eada.2020.1.04.

[39] C. Wang, C. Deng, and S. Wang, "Imbalance-XGBoost: Leveraging weighted and focal losses for binary label-imbalanced classification with XGBoost," *Pattern Recognit. Lett.*, vol. 136, pp. 190–197, Aug. 2020, doi: 10.1016/j.patrec.2020.05.035.

[40] S. Ryou, S.-G. Jeong, and P. Perona, "Anchor loss: Modulating loss scale based on prediction difficulty," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 5991–6000.

[41] H. Yu, C. Sun, X. Yang, S. Zheng, Q. Wang, and X. Xi, "LW-ELM: A fast and flexible cost-sensitive learning framework for classifying imbalanced data," *IEEE Access*, vol. 6, pp. 28488–28500, 2018, doi: 10.1109/ACCESS.2018.2839340.

[42] R. Daneshjou, C. Barata, B. Betz-Stablein, M. E. Celebi, N. Codella, M. Combalia, P. Guitera, D. Gutman, A. Halpern, B. Helba, H. Kittler, K. Kose, K. Liopyris, J. Malvehy, H. S. Seog, H. P. Soyer, E. R. Tkaczyk, P. Tschandl, and V. Rotemberg, "Checklist for evaluation of image-based artificial intelligence reports in dermatology: CLEAR derm consensus guidelines from the international skin imaging collaboration artificial intelligence working group," *JAMA Dermatol.*, vol. 158, no. 1, p. 90, Jan. 2022, doi: 10.1001/jamadermatol.2021.4915.

[43] J. Höhn, A. Hekler, J. S. Utikal, N. Grabe, D. Schadendorf, J. Klode, C. Berking, T. Steeb, A. H. Enk, and C. Von Kalle, "Skin cancer classification using convolutional neural networks with integrated patient data: A systematic review," *J. Med. Internet Res.*, vol. 23, no. 7, May 2020, Art. no. e20708, doi: 10.2196/20708.

[44] C. Turkay, A. Lundervold, A. J. Lundervold, and H. Hauser, "Hypothesis generation by interactive visual exploration of heterogeneous medical data," in *Proc. Int. Workshop Hum.-Comput. Interact. Knowl. Discovery Complex, Unstructured, Big Data*, in Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 7947, 2013, pp. 1–12, doi: 10.1007/978-3-642-39146-0_1.

[45] L. Yue, D. Tian, W. Chen, X. Han, and M. Yin, "Deep learning for heterogeneous medical data analysis," *World Wide Web*, vol. 23, no. 5, pp. 2715–2737, Sep. 2020, doi: 10.1007/S11280-019-00764-Z.

[46] K. J. Cios and G. William Moore, "Uniqueness of medical data mining," *Artif. Intell. Med.*, vol. 26, nos. 1–2, pp. 1–24, Sep. 2002, doi: 10.1016/S0933-3657(02)00049-0.

[47] N. Lama, R. Kasmi, J. R. Hagerty, R. J. Stanley, R. Young, J. Miinch, J. Nepal, A. Nambisan, and W. V. Stoecker, "ChimeraNet: U-Net for hair detection in dermoscopic skin lesion images," *J. Digit. Imag.*, vol. 36, no. 2, pp. 526–535, Nov. 2022, doi: 10.1007/s10278-022-00740-6.

[48] Q. Abbas, M. E. Celebi, and I. F. García, "Hair removal methods: A comparative study for dermoscopy images," *Biomed. Signal Process. Control*, vol. 6, no. 4, pp. 395–404, Oct. 2011, doi: 10.1016/j.bspc.2011.01.003.

[49] P. A. Lyakhov, U. A. Lyakhova, and N. N. Nagornov, "System for the recognizing of pigmented skin lesions with fusion and analysis of heterogeneous data based on a multimodal neural network," *Cancers*, vol. 14, no. 7, p. 1819, Apr. 2022, doi: 10.3390/cancers14071819.

[50] P. Cerda, G. Varoquaux, and B. Kégl, "Similarity encoding for learning with dirty categorical variables," *Mach. Learn.*, vol. 107, pp. 1477–1494, Sep. 2018.

[51] T. Al-Shehari and R. A. Alsowail, "An insider data leakage detection using one-hot encoding, synthetic minority oversampling and machine learning techniques," *Entropy*, vol. 23, no. 10, p. 1258, Sep. 2021, doi: 10.3390/e23101258.

[52] P. Rodríguez, M. A. Bautista, J. Gonzàlez, and S. Escalera, "Beyond one-hot encoding: Lower dimensional target embedding," *Image Vis. Comput.*, vol. 75, pp. 21–31, Jul. 2018, doi: 10.1016/j.imavis.2018.04.004.

[53] K. Potdar, "A comparative study of categorical variable encoding techniques for neural network classifiers," *Int. J. Comput. Appl.*, vol. 175, no. 4, pp. 7–9, Oct. 2017, doi: 10.5120/ijca2017915495.

[54] Q. Wang, Y. Ma, K. Zhao, and Y. Tian, "A comprehensive survey of loss functions in machine learning," *Ann. Data Sci.*, vol. 9, no. 2, pp. 187–212, Apr. 2022, doi: 10.1007/S40745-020-00253-5.

[55] Y. Kim, Y. Lee, and M. Jeon, "Imbalanced image classification with complement cross entropy," *Pattern Recognit. Lett.*, vol. 151, pp. 33–40, Nov. 2021, doi: 10.1016/j.patrec.2021.07.017.

[56] Y. Ho and S. Wookey, "The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling," *IEEE Access*, vol. 8, pp. 4806–4813, 2020, doi: 10.1109/ACCESS.2019.2962617.

[57] Q. Jodelet, X. Liu, and T. Murata, "Balanced softmax cross-entropy for incremental learning," in *Proc. Int. Conf. Artif. Neural Netw.*, in Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 12892, 2021, pp. 385–396, doi: 10.1007/978-3-030-86340-1_31.

[58] E. Tasci, Y. Zhuge, K. Camphausen, and A. V. Krauze, "Bias and class imbalance in oncologic data—Towards inclusive and transferrable AI in large scale oncology data sets," *Cancers*, vol. 14, no. 12, p. 2897, Jun. 2022, doi: 10.3390/cancers14122897.

[59] T. Huynh, A. Nibali, and Z. He, "Semi-supervised learning for medical image classification using imbalanced training data," *Comput. Methods Programs Biomed.*, vol. 216, Apr. 2022, Art. no. 106628, doi: 10.1016/j.cmpb.2022.106628.

[60] A. C. Foahom Gouabou, R. Iguernaissi, J.-L. Damoiseaux, A. Moudafi, and D. Merad, "End-to-end decoupled training: A robust deep learning method for long-tailed classification of dermoscopic images for skin lesion classification," *Electronics*, vol. 11, no. 20, p. 3275, Oct. 2022, doi: 10.3390/electronics11203275.

[61] N. H. Vo and Y. Won, "Classification of unbalanced medical data with weighted regularized least squares," in *Proc. Frontiers Converg. Biosci. Inf. Technol.*, Mar. 2007, pp. 347–352, doi: 10.1109/FBIT.2007.20.

[62] Y. S. Aurelio, G. M. de Almeida, C. L. de Castro, and A. P. Braga, "Learning from imbalanced data sets with weighted cross-entropy function," *Neural Process. Lett.*, vol. 50, no. 2, pp. 1937–1949, Oct. 2019, doi: 10.1007/s11063-018-09977-1.

[63] Y. Dong, X. Shen, Z. Jiang, and H. Wang, "Recognition of imbalanced underwater acoustic datasets with exponentially weighted cross-entropy loss," *Appl. Acoust.*, vol. 174, Mar. 2021, Art. no. 107740, doi: 10.1016/j.apacoust.2020.107740.

[64] S. Wang, Y. Yin, D. Wang, Y. Wang, and Y. Jin, "Interpretability-based multimodal convolutional neural networks for skin lesion diagnosis," *IEEE Trans. Cybern.*, vol. 52, no. 12, pp. 12623–12637, Dec. 2022, doi: 10.1109/TCYB.2021.3069920.

[65] G. Goh, N. Cammarata, C. Voss, S. Carter, M. Petrov, L. Schubert, A. Radford, and C. Olah, "Multimodal neurons in artificial neural networks," *Distill*, vol. 6, no. 3, pp. 1–12, Mar. 2021, doi: 10.23915/distill.00030.

[66] K. Liu, Y. Li, N. Xu, and P. Natarajan, "Learn to combine modalities in multimodal deep learning," 2018, *arXiv:1805.11730*.

[67] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," 2015, *arXiv:1511.08458*.

[68] J. Lyu, H. Shi, J. Zhang, and J. Norvilitis, "Prediction model for suicide based on back propagation neural network and multilayer perceptron," *Frontiers Neuroinform.*, vol. 16, p. 79, Aug. 2022, doi: 10.3389/fninf.2022.961588.

[69] E. K. Y. Yapp, X. Li, W. F. Lu, and P. S. Tan, "Comparison of base classifiers for multi-label learning," *Neurocomputing*, vol. 394, pp. 51–60, Jun. 2020, doi: 10.1016/j.neucom.2020.01.102.

[70] *ISIC Archive*. Accessed: Feb. 7, 2023. [Online]. Available: https://www.isic-archive.com/#!/topWithHeader/onlyHeaderTop/gallery?filter=%5B%5D

[71] *The ISIC 2020 Challenge Dataset | ISIC 2020 Challenge Dataset*. Accessed: Oct. 29, 2023. [Online]. Available: https://challenge2020.isic-archive.com/

[72] B. Cassidy, C. Kendrick, A. Brodzicki, J. Jaworek-Korjakowska, and M. H. Yap, "Analysis of the ISIC image datasets: Usage, benchmarks and recommendations," *Med. Image Anal.*, vol. 75, Jan. 2022, Art. no. 102305, doi: 10.1016/j.media.2021.102305.

[73] J. A. Siegel, K. Korgavkar, and M. A. Weinstock, "Current perspective on actinic keratosis: A review," *Brit. J. Dermatol.*, vol. 177, no. 2, pp. 350–358, Aug. 2017, doi: 10.1111/bjd.14852.

[74] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269. Accessed: Dec. 30, 2022.

[75] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-ResNet and the impact of residual connections on learning," in *Proc. Int. Conf. Learn. Represent. Workshop*, Feb. 2016, pp. 4278–4284.

[76] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5987–5995.

[77] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie, and L. Farhan, "Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions," *J. Big Data*, vol. 8, no. 1, pp. 1–11, Mar. 2021, doi: 10.1186/s40537-021-00444-8.

[78] *Lyakhova-Uliana/Multimodal-Analysis-of-Unbalanced-Dermatological-Data-for-Skin-Cancer-Recognition*. Accessed: Sep. 24, 2023. [Online]. Available: https://github.com/Lyakhova-Uliana/Multimodal-Analysis-of-Unbalanced-Dermatological-Data-for-Skin-Cancer-Recognition

[79] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, no. 1, pp. 1–13, 2020, doi: 10.1186/S12864-019-6413-7.

[80] J. Yap, W. Yolland, and P. Tschandl, "Multimodal skin lesion classification using deep learning," *Experim. Dermatol.*, vol. 27, no. 11, pp. 1261–1267, Nov. 2018, doi: 10.1111/exd.13777.

[81] X. He, Y. Wang, S. Zhao, and X. Chen, "Co-attention fusion network for multimodal skin cancer diagnosis," *Pattern Recognit.*, vol. 133, Jan. 2023, Art. no. 108990, doi: 10.1016/j.patcog.2022.108990.

[82] A. G. C. Pacheco and R. A. Krohling, "The impact of patient clinical information on automated skin cancer detection," *Comput. Biol. Med.*, vol. 116, Jan. 2020, Art. no. 103545, doi: 10.1016/j.compbiomed.2019.103545.

[83] K. Sriwong, S. Bunrit, K. Kerdprasop, and N. Kerdprasop, "Dermatological classification using deep learning of skin image and patient background knowledge," *Int. J. Mach. Learn. Comput.*, vol. 9, no. 6, pp. 862–867, Dec. 2019, doi: 10.18178/ijmlc.2019.9.6.884.

[84] Q. Chen, M. Li, C. Chen, P. Zhou, X. Lv, and C. Chen, "MDFNet: Application of multimodal fusion method based on skin image and clinical data to skin cancer classification," *J. Cancer Res. Clin. Oncol.*, vol. 149, pp. 1–13, Aug. 2022, doi: 10.1007/S00432-022-04180-1.

[85] M. Castro-Fernández, A. Hernández, H. Fabelo, F. J. Balea-Fernández, S. Ortega, and G. M. Callicó, "Towards skin cancer self-monitoring through an optimized MobileNet with coordinate attention," in *Proc. 25th Euromicro Conf. Digit. Syst. Design (DSD)*, Aug. 2022, pp. 607–614, doi: 10.1109/DSD57027.2022.00087.

[86] Q. Ha, B. Liu, and F. Liu, "Identifying melanoma images using EfficientNet ensemble: Winning solution to the SIIM-ISIC melanoma classification challenge," 2020, *arXiv:2010.05351*.

[87] R. H. Hridoy, F. Akter, and A. Rakshit, "Computer vision based skin disorder recognition using EfficientNet: A transfer learning approach," in *Proc. Int. Conf. Inf. Technol. (ICIT)*, Jul. 2021, pp. 482–487, doi: 10.1109/ICIT52682.2021.9491776.

[88] T. J. Brinker et al., "Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task," *Eur. J. Cancer*, vol. 113, pp. 47–54, May 2019, doi: 10.1016/j.ejca.2019.04.001.

[89] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, Feb. 2017, doi: 10.1038/nature21056.

[90] T. J. Brinker, A. Hekler, A. H. Enk, C. Berking, S. Haferkamp, A. Hauschild, M. Weichenthal, J. Klode, D. Schadendorf, T. Holland-Letz, C. von Kalle, S. Fröhling, B. Schilling, and J. S. Utikal, "Deep neural networks are superior to dermatologists in melanoma image classification," *Eur. J. Cancer*, vol. 119, pp. 11–17, Sep. 2019, doi: 10.1016/j.ejca.2019.05.023.

[91] K. Fogelberg, S. Chamarthi, R. C. Maron, J. Niebling, and T. J. Brinker, "Domain shifts in dermoscopic skin cancer datasets: Evaluation of essential limitations for clinical translation," *New Biotechnol.*, vol. 76, pp. 106–117, Sep. 2023, doi: 10.1016/j.nbt.2023.04.006.

PAVEL A. LYAKHOV was born in 1988. He received the degree in mathematics from Stavropol State University, in 2009, and the Ph.D. degree in physical and mathematical sciences. He is the Head of the Department of Mathematical Modeling, North-Caucasus Federal University; and the Department of Modular Computing and Artificial Intelligence, North-Caucasus Center for Mathematical Research (regional scientific and educational mathematical center), North-Caucasus Federal University. He is the author of more than 200 publications. His research interests include digital signal and image processing, artificial intelligence, neural networks, modular arithmetic, parallel computing, high-performance computing, digital circuits, and hardware accelerators.

ULYANA A. LYAKHOVA was born in 1997. She received the bachelor's degree in biology and the master's degree in applied mathematics and informatics from North-Caucasus Federal University, in 2018 and 2020, respectively. She is a Postgraduate Student and a Junior Researcher with the Department of Mathematical Modeling, Faculty of Mathematics and Computer Science named after Prof. N. I. Chervyakov North-Caucasus Federal University. She is also a Junior Researcher with the Department of Modular Computing and Artificial Intelligence, North-Caucasus Center for Mathematical Research (regional scientific and educational mathematical center), North-Caucasus Federal University. She is the author of more than 20 scientific papers. Her research interests include digital processing of visual data, artificial intelligence, and medical data processing.

DIANA I. KALITA was born in 1990. She received the degree in applied mathematics and computer science from North-Caucasus Federal University, in 2013, and the Ph.D. degree in computer sciences. She is an Associate Professor and a Junior Researcher with the Department of Mathematical Modeling, North-Caucasus Federal University. She is the author of more than 30 publications. Her research interests include digital image processing, modular arithmetic, parallel computing, high-performance computing, digital circuits, and hardware accelerators.