Ministry of Science and Higher Education of the Republic of Kazakhstan
SDU University

Daniyar Kuanyshbayev

A thorough survey into the recognition of face emotion expression:experimental study, practical uses, and recommendations for the future

THESIS

Presented in Partial Fulfilment for the

Degree of Master in Computer Science

(degree code: 7M06102)

Department of Computer Science

Faculty of Engineering and Natural Sciences

Supervisor: **Meraryslan Meraliyev**

Kaskelen, June 2024

SDU University
Faculty of Engineering and Natural Sciences
Department of Computer Science

Dean of Faculty of Engineering and Natural Sciences

Assistant Professor, Ph.D. Akhmedov Ramis

____

« ______» ______________________________ 2024

Topic of the thesis:

A thorough survey into the recognition of face emotion expression:experimental study, practical uses, and recommendations for the future

Thesis submitted as part of the requirements for the award of the MSc in "7M06102 - Computer Science", SDU University

Head of Department Zhanar Mukash

Academic Supervisor Meraryslan Meraliyev

Master student Daniyar Kuanyshbayev

Kaskelen, 2024

Declaration

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Daniyar Kuanyshbayev

June 2024

Abstract

The growth of the volume of information, as well as the expansion of the range of technically complex decision-making tasks require the systematization of existing methods and the development of new techniques and algorithms for their solution. The master's thesis examines the possibility of using a neural network to solve the problem of recognizing human emotions. Artificial neural networks offer promising prospects for development, and software has a great advantage in using them. Moreover, each task performed has an unlimited and non-standard set of solution methods. The article considers the possibility of using a neural network to solve the problem of recognizing human emotions. The increasing volume of data, along with the breadth of technologically sophisticated issues with solving, necessitates the systematization of existing approaches and the creation of new techniques and algorithms for their resolution. The master's thesis investigates the feasibility of utilizing a neural network to tackle the challenge of identifying human emotions. Artificial neural networks provide tremendous growth opportunities, and software can benefit greatly from their use. Furthermore, each challenge contains an infinite and non-standardized collection of solution techniques. The article discusses the feasibility of utilizing a neural network to tackle the difficulty of identifying human emotions.

Аңдатпа

Ақпарат көлемінің өсуі, сондай-ақ шешім қабылдаудың техникалық күрделі мәселелерінің спектрін кеңейту қолданыстағы әдістерді жүйелеуді және оларды шешудің жаңа әдістері мен алгоритмдерін әзірлеуді талап етеді. Магистрлік диссертация адамның эмоциясын тану мәселесін шешу үшін нейрондық желіні пайдалану мүмкіндігін қарастырады. Жасанды нейрондық желілер дамудың перспективаларын ұсынады, ал бағдарламалық жасақтама оларды қолдануда үлкен артықшылыққа ие. Сонымен қатар, орындалатын әрбір тапсырмада шешудің шексіз және стандартты емес әдістері бар. Мақалада адамның эмоциясын тану мәселесін шешу үшін нейрондық желіні пайдалану мүмкіндігі қарастырылады.

Аннотация

Растущий объем данных, а также широта технологичности решаемых задач обуславливают необходимость систематизации существующих подходов и создания новых методик и алгоритмов для их решения. В магистерской диссертации исследуется возможность использования нейронной сети для решения задачи идентификации человеческих эмоций. Искусственные нейронные сети предоставляют огромные возможности для развития, и программное обеспечение может извлечь большую пользу из их использования. Кроме того, каждая задача содержит бесконечную и нестандартизированную коллекцию методов решения. Эта статья рассматривается возможность использования нейронной сети для решения задачи идентификации человеческих эмоций.

Abbreviations

ANN – Artificial Neural Networks

AIS – Artificial Intelligent Systems

SVM – Support Vector Machines

SCLD – Scale of Consciousness of the Level of Development

SKLiD – Facial Movement Coding System

GA – Genetic Algorithms

LVQ – Vector Quantization (Kohonen Network)

AAM – Active Appearance Model

ASM – Active Shape Model

EBGM – Elastic Bunch Graph Matching

PCA – Principal Component Analysis

LDA – Linear Discriminant Analysis

LBP – Local Binary Patterns

HoG – Histogram of Oriented Gradients

Table of Contents

Introduction

In the modern world, artificial intelligence systems have gained great popularity due to their rich capabilities and efficiency of use.

There are many tasks that currently require practical solutions using intelligent systems: economics and business, robotics, exploration, mathematics, biophysics, avionics, safety and security systems, medicine and many others. Such an abundance of applications suggests that intelligent systems are a unique set for solving problems of analyzing and processing large amounts of data, solving problems of varying degrees of complexity.

Neural networks made up of are now widely employed in artificial intelligence systems.

Neural network is a computational structure consisting of many elements of the same type. These elements perform simple functions, and all processes occurring in an artificial neural network can be related to processes occurring in the nervous system of living organisms.

Neural networks are nonlinear in nature, they do not have an explicit dependence, which allows you to immediately use the developed technology (information model of a neural network). Linear modeling has been the main modeling method for many years, since optimization procedures have been well developed for it.

The importance of the master's thesis lies in the use of neural networks to solve poorly formalized data mining problems. The increasing volume of data, in addition to a wider array of theoretically complicated making decisions tasks, need the organization of current approaches and a generation of new strategies and algorithms for their resolution. The master's thesis investigates the feasibility of utilizing artificial neural networks to tackle the challenge of identifying human emotions.

The creation of artificial neural networks presents exciting opportunities, and software greatly benefits from their use. Furthermore, there is an infinite and non-standard collection of solution techniques for every task completed. The master's thesis investigates the viability of applying neural networks to the challenge of emotion recognition in humans.

Purpose of the work is to **select an information model of a neural network and implement an algorithm for recognizing facial motor activity. A neural network should be optimal in terms of its internal structure, a way to control information flows between neurons. The selected information model will be used to solve a practical problem**.

The main tasks of the dissertation work:

1) The study of existing types of artificial intelligence systems, as well as meth-

ods of their functioning.

2) The study of the main types of information models of artificial neural networks. Choosing the optimal information model of a neural network for solving the problem of emotion recognition.

3) The study of existing methods of facial recognition and the identification of universal methods among them.

4) Implementation and description of an algorithm for recognizing facial motor activity for an intelligent system and solving a practical problem.

The subject of this research is approaches and methods of facial expression recognition.

The object of the study is information models of artificial neural networks, as well as the implementation and description of the facial motor activity recognition algorithm.

The scientific novelty of the master's thesis is the use of neural network technologies (information models), as well as a system of facial motor activity to implement an algorithm for recognizing human emotions.

To implement a neural network that will recognize emotions, it is necessary to identify key facial features and key facial expressions that distinguish a particular person from many others.

The complexity of the implementation lies in the training of a neural network. The choice of source data for the task is primarily a mathematical and geometric description of facial expressions.

The automated processing and analysis of visual information is becoming a more pressing issue due to the growing complexity of the scientific and technical problems being handled. Digital and smart cameras already exist, as well as image processing software.

Modern technologies are rapidly moving towards the moment when computers will become independent systems capable of self-learning and the growth of "personality", as it happens to a person from the moment of his birth. Many developers are already solving complex problems using computer vision machines, using various methods of operation of the system, including neural networks.

The novelty of the master's thesis lies in the development and research of face and emotion recognition algorithms for their further application in web applications. As part of the work, an integrated system has been created that uses advanced artificial neural network techniques to analyze and interpret facial motor activity, which allows for accurate and efficient identification of human emotions. This system is designed to be implemented in interactive web applications, which makes it particularly relevant for areas where a high level of user interaction is required, such as online learning, telemedicine and customer service. The considered algorithms take into account not only technical aspects of neural networks, but also practical applicability in modern web technologies, which makes the study significant from both scientific and applied points of view.

The dissertation work consists of an introduction, three sections, a conclusion and a list of references.

The first part examines theoretical materials on neural networks, as well as existing types of artificial intelligence systems and methods for recognizing emotions.

The second part contains a detailed description and layout of information models of neural networks, their characteristics, comparative analysis and selection of the optimal model for solving a practical problem. There is also an analysis of methods for determining faces in an image and a set of logical rules for recognizing emotions (a system for encoding facial movements according to P. Ekman).

The third part presents an algorithm for recognizing emotions using a neural network.

Chapter 1

Basic Concepts of Artificial Intelligent Systems and Neural Networks

1.1 The Development of Neural Networks

Artificial neural networks were first considered in 1940s.as In the scientific field, the theory of neural networks was reflected in the work of McCulloch and Pitts in 1943 [1]. In this article, it was argued that almost any logical or arithmetic function can be implemented using the simplest neural network.

Among the fundamental works, it should be noted the model of D. Hebb, who formulated the law in 1949, which became the starting point for training neural networks. Hubb was the first to suggest that learning consists primarily in changing the strength of synoptic connections. Hebb's theory is a typical example of self-learning, when the system under test is trained to perform the required task without the intervention of an experimenter. M. Minsky also contributed through the study of many problems, including the well-known problem of ”exclusion” [1].

F. In 1958, Rosenblatt proposed a neural network called perceptron, which was designed to classify objects. During the training, perceptron received a message from a ”teacher”. Thanks to the message, you can determine which class this object belongs to. In addition, the trained perceptron was able to sort previously unused objects on its own, while making a small number of mistakes.

The period of lull in the development of neural networks occurred in 1968-1985. With the advent of high-performance personal computers, this has become available for modeling neural networks.

The work of F. It has become a reference book for specialists interested in the theory of neural networks ”Neurocomputer technologies”.

After the appearance of D. Hopfield's work (1982), interest in neural networks increased dramatically. Based on Hubb's teaching rules, Hopfield showed that tasks with neurons can be reduced to generalization of many models developed by that time in the physics of disordered systems.

Then, in 1986, the work of Williams, Rumelhart and Hinton appeared, which answered an important question about behavior in learning multilayer neural networks. After that, the algorithm proposed by Hinton underwent a number of changes [2].

In the 80s, a solid theoretical foundation was gradually formed, on the basis of which most networks are being created today. The developed theory has been widely used over the past two decades to solve applied problems. Companies began to appear that develop software for building artificial neural networks. In the 90s, neural networks began to be used in business, where they showed tremendous effectiveness in solving many tasks - from forecasting demand for products to analyzing the solvency of bank customers.

In 2007, he created deep learning algorithms for neural networks at the University of Toronto. When training the lower levels of the network, Hinton used a limited Boltzmann machine representing a stochastic repeating neural network. After exploring the network, the resulting application could quickly solve the task (for example, find faces in a photo). Currently, this feature is built into all digital cameras. A similar technology is used by Internet search engines when sorting images [3].

The automated processing and analysis of visual information is becoming a more pressing issue due to the growing complexity of the scientific and technical problems being handled.

Despite the fact that the study of neural modeling has been conducted for more than sixty years, there is not a single area of the brain where the information processing process would be completely understandable. Also, there is not a single neuron for which it would be possible to define a code for transmitting information in the form of a sequence of pulses.

Currently, there are a large number of neural network configurations that differ in the principles of operation, and, therefore, are designed for different tasks.

The future of neurocomputer technologies will be associated with new discoveries in the field of neural modeling – as soon as it becomes possible to solve the mystery of the functioning of at least one area of the brain, much will immediately become clear about its other areas.

1.2 The concept of ≪Artificial neural network≫

An artificial neural network is a mathematical model that is based on the principles of biological neural networks, which are the nerve cells of a live creature. It may be implemented both software- and hardware-wise. This idea came about as an attempt to mimic the functions that take place in the human brain [4].

A system of basic processors, or artificial neurons, linked and interacting with one another is called an artificial neural network. Periodically, signals are received by or sent to other processors by each of the network processors. The most challenging jobs can be completed by a big network in the smallest amount of time.

Neural networks serve as a theoretical method for solving nonlinear optimization issues. Neural network theory is applied in cybernetics to solve adaptive control issues and develop robotics algorithms [5].

In programming, a neural network is one of the ways to solve the problem of effective parallelism.

Instead of creating code, neural network programming entails learning about the network. The network may learn to recognize relationships between input and output data, generalize, simplify findings, and utilize its expertise to decompose difficult jobs into easier ones.

1.2.1 Biological neuron and artificial neuron

The human brain and its nervous system consist of neurons connected by nerve fibers. Electrical impulses are transmitted between neurons using nerve fibers. All actions that occur with a living organism, all irritations of the skin, eyes, pain, thinking processes – there is an interaction between neurons. The structure of a biological neuron is shown in Figure 1:

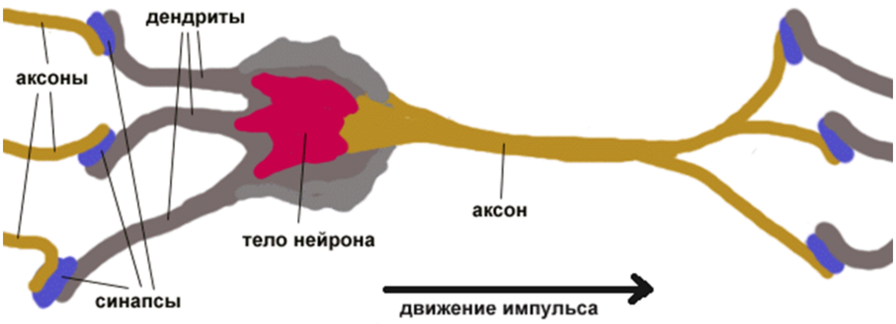

Figure 1.1 – Biological neuron

image was taken from - 19

Dendrites - receives the impulses of a neuron;

Axon - transmits the impulse of a neuron;

Synapses are formations that affect the strength of the impulse, for the contact of the axon and the dendrite.

During the passage of the synapse, the strength of the pulse changes a certain number of times (the weight of the synapse). When impulses arrive at a neuron along several dendrites, they are summed up. If the total pulse exceeds the threshold, then the neuron goes into a state of excitation, forms its own impulse and sends it further along the axon. The behavior of the corresponding neuron can change, since the weights of synapses tend to change over time. The mathematical model of the described process is presented as follows (Figure 2):

This model describes a neuron with three inputs (dendrites), where synapses have weights $_1$, $_2$, $_3$, to whom the forces are coming $_1$, $_2$, $_3$ accordingly. The neuron receives impulses $_{11}$, $_{22}$, $_{33}$ after passing synapses and dendrites. The resulting total impulse x=$_{11}$+$_{22}$+$_{33}$ the neuron transforms according to the transfer function f(x). y = f(x)=f($_{11}$+$_{22}$+$_{33}$) – the strength of the output pulse. To summarize, we get a set of numbers (vector) in the form of inputs. Next, the neuron outputs a certain number of y at the output.

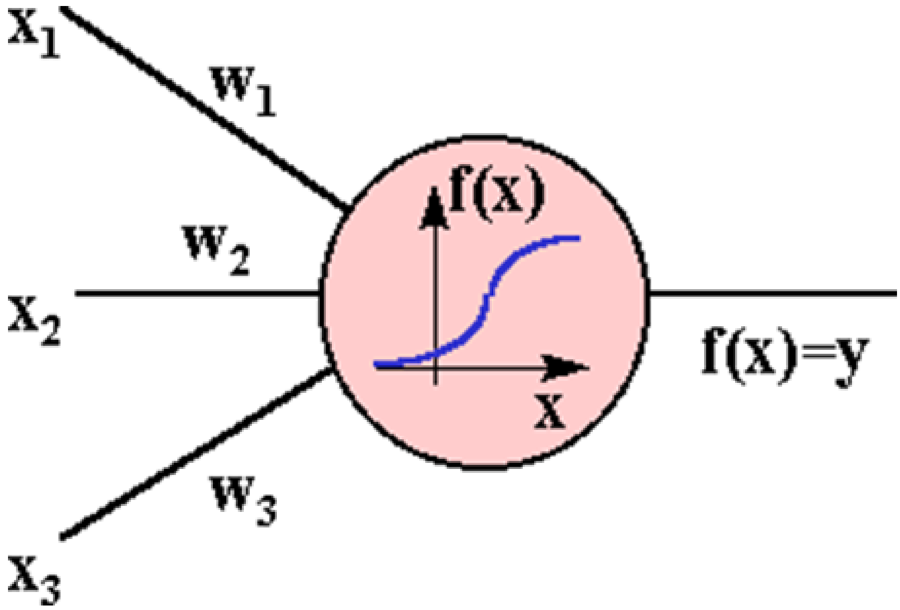

Figure 1.2 – Mathematical model of a neuron

An artificial neuron looks like this: its input receives many signals, each of which is simultaneously the output of another neuron. Such an input is multiplied by the corresponding weight, then the products are summed, determining the level of activation of the neuron.

The model implementing this idea is shown in Figure 3:

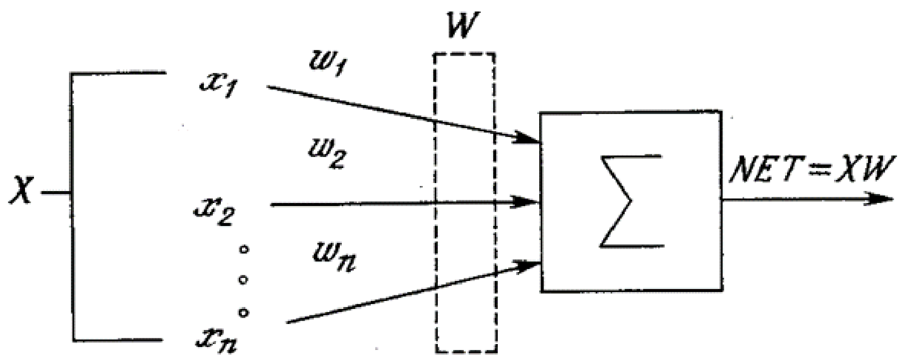

Figure 1.3 – Neuron activation model

Thus, a group of signals x_1, x_2, ...,x_n, enters the artificial neuron's input, collectively they are represented via vector X. These signals are similar to those received at the inputs of a biological neuron. Additionally, a signal is multiplied by the appropriate weight. w_1, w_2,..., w_n, then summed in the summing block $\sum$. Each weight is equal to the strength of one synaptic connection in a biological neuron. The output, which is usually called NET, is created from a summing block where the weighted elements are added algebraically.

Converting the NET signal involves a conventional linear function called activation. It is denoted by F and gives an output signal OUT.

OUT=K(NET), where K is the constant of the OUT=1 threshold function if NET>T OUT = 0 in all different cases, T is a a continuous threshold value that better simulates a neural network.

A synthetic neuron exhibiting an activation function is seen in Figure 4:

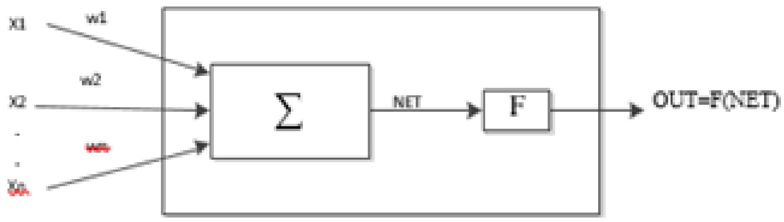

Figure 1.4 – Activation function of a neuron

The block marked F receives the NET signal and outputs the OUT signal.

F is called a compressive function if, for any values of NET , the values of OUT belong to some finite interval [6].

The artificial neuron model ignores most of the properties of a biological neuron. For example, time delays that affect the dynamics of the system.

An output signal is produced instantaneously by the input signals. Furthermore, the impacts of the real neuron's synchronizing role are ignored by the artificial neuron.

However, it should be noted the exceptional similarity of a living neuron and an artificial one.

To determine the place of neural networks in the field of information technology, it is necessary to refer to the classification of artificial intelligent systems.

1.3 Classifications of artificial intelligent systems

1.3 Classifications of artificial intelligent systems

Artificial intelligent systems have characteristic features:

Firstly, developed communication skills that characterize the way the user's computer interacts with the system. The possibility of contacting the system with an arbitrary request in a dialogue with an intelligent system is not excluded. At the same time, the language of the intellectual system should be as close as possible to the natural language.

Secondly, the solution of poorly formalized tasks, i.e. tasks that do not have a specific solution, but require a non-standard approach, depending on the situation, existing data and the final result. Poorly formalized tasks are effectively solved using artificial neural networks.

Thirdly, the ability to self–study - that is, the ability of an intellectual system to extract knowledge from the accumulated experience of specific situations. Pre-training of the system requires processed initial data.

In accordance with the presented features, intelligent systems can be divided into the following (Table 1):

Table 1.1 – Types of artificial intelligent systems

A type of artificial intelligent system	Type of artificial intelligent system
Systems with switching capabilities	-intelligent databases; -natural language interfaces; - hypertext systems; -contextual help systems; -cognitive graphics.
Expert systems	- classification systems; - additional defining systems; - transformative systems; - multi-antenna systems.
Self-learning systems	- inductive systems; - neural networks; - systems based on precedents; - information storages.
Adaptive systems	- CASE technologies; - component technology.

Smart data bases are different from traditional databases in that they may get the required data upon request; this data may not be explicitly kept but is instead produced from the database.

Naturally, the language interface transforms natural language constructs into an intra-machine level of knowledge representation. It is utilized for voice input instructions in control systems, machine translation from other languages, contextually examination of documented text material, and access to smart databases..

Hypertext systems are used in databases of textual information, where a keyword search is required, and have a more complex semantic organization of keywords.

In contextual help systems, the user describes the problem (situation), and the system uses an additional dialog to specify it and search for recommendations suitable for this situation. Such systems are created as an application to documentation systems and belong to the class of knowledge dissemination systems.

Operational processes are monitored and managed with the help of cognitive graphics technologies. Numerous aspects of the topic under study are described using graphic pictures in a visual and integrated style.

Expert systems are made to resolve issues using a body of information that has been acquired over time and represents the expertise of subject matter specialists in the issue at hand.

Multi-agent systems are dynamic systems that exchange findings dynamically by integrating several heterogeneous information sources into the knowledge base.

The foundation of self-learning systems is automated classification of real-world scenario samples [6].

Characteristic features of self-learning systems are:

- "with a teacher" self-learning systems, in which the value of each example's class-forming feature—the trait that indicates that it belongs to a certain class of situations—is explicitly specified;

- "without a teacher" self-learning systems, in which the system recognizes classes of circumstances based on how close the values of the categorization characteristics are to one another.

Inductive systems generalize examples according to the principle from the particular to the general, and the generalization process is carried out as follows:

1) From a list of predetermined features, one is chosen for categorization (either in a rule-based or sequential manner).

2) Many instances are split into subgroups based on the value of the chosen property.

3) The example is examined to see if it is a member of the same class.

4) The classification process is completed (the other classification characteristics are not taken into consideration) if a subset of instances falls into a single subclass, meaning that every example in the subset has the same value of the class-forming feature.

5) The classification procedure is repeated, commencing at point 1, for subsets of samples that have a mismatched value of the class-forming characteristic.

Neural networks are tools for parallel computing that consist of many simple processors that periodically receive and send signals to other processors.

Next, we should turn to intelligent information systems from the point of view of the tasks to be solved. The following systems can be distinguished here:

- management systems and reference systems;
- computational linguistics systems;
- recognition systems;
- game systems;
- systems for creating intelligent information systems.

Figure 5 shows the classification of intelligent information systems according to the tasks to be solved:

Systems can solve not one, but several tasks, or solve a number of other tasks in the process of solving one task.

Intelligent systems are also classified according to such criteria as "methods used". Soft, hard and hybrid methods can be distinguished here (Figure 6):

A sophisticated computer technology known as "soft computing" is based on probabilistic computing, fuzzy logic, genetic computing, and neurocomputing.

Standard computer computing, or "hard computing," is unrelated to "soft computing."

Systems that use several computer technologies—or, in the case of intelligent systems, artificial intelligence technologies—are called hybrid systems [7].

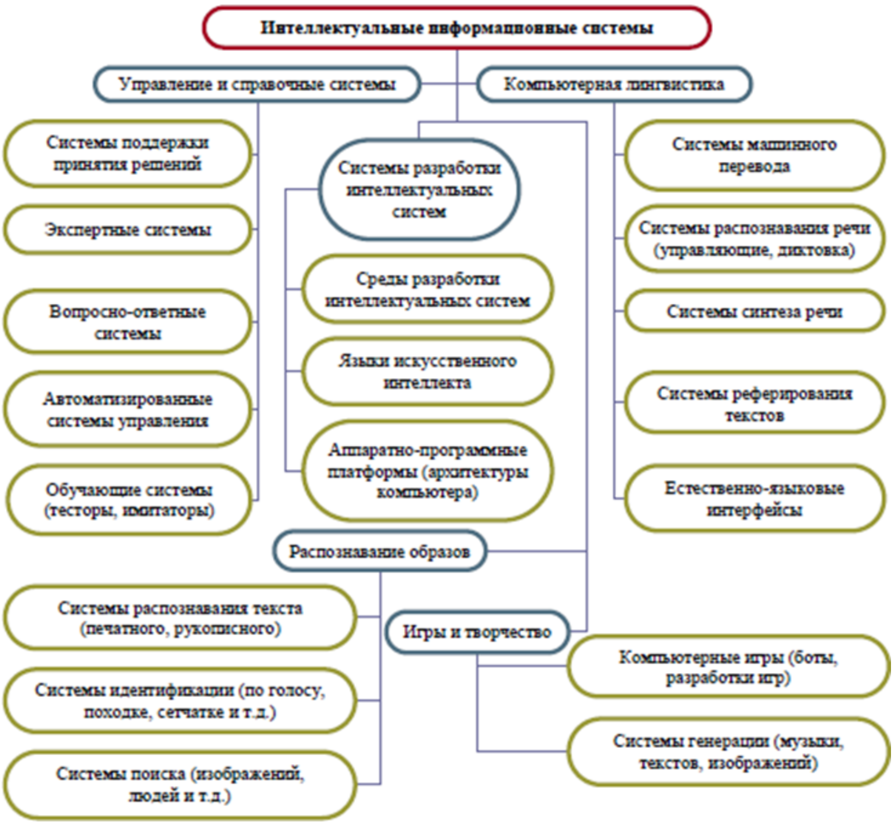

Figure 1.5 – Classification of IP by tasks to be solved

A vast body of theoretical information from several scientific disciplines is covered in the categorization of artificially intelligent systems. An AIS's development and deployment are difficult processes from the start to the finish.

When developing an intelligent emotion recognition system, it should be said about the most popular methods used in psychology, criminology and other fields.

1.4 Profiling in a data mining system

At the initial stage of its development, the term "profiling" meant the compilation of a psychological portrait (profile) of a criminal in the footsteps at the crime scene. Otherwise, it's called personality profiling. First of all, the method is used in the search for serial killers. It combines knowledge of criminology, psychiatry and psychology. Later, profiling began to be called the identification of potentially dangerous individuals. At first, the technology was used in civil aviation during the pre-flight inspection of passengers.

A collection of psychological tools and procedures known as profiling are used to assess and forecast human behavior by analyzing the most telling personal indicators, physical attributes, and verbal and nonverbal cues [8].

The technique is used by expert profilers, i.e. specialists in the field of lie recognition. Profiling theory can be effectively applied in computer vision based on classification according to the identified signs for the potential danger of the

Figure 1.6 – Classification of IP by methods

subject or recognition of the criminal's lies.

To describe emotions, it is necessary to refer to the facial movement coding system (abbreviated SKLiD) developed by Paul Ekman and Wallace Friesen in 1978 [9].

Using SKLIDS, you can encode a facial expression, creating a face model from certain units of action, as well as a fixed period of time to reproduce a particular facial expression. In the SKLiD manual, Paul Ekman interprets emotions by contraction or tension of one or more muscle groups.

The encoding system has fixed descriptor numbers that have significant differences with the units of action. If the units of action are precisely manifested in behavior, then the muscle base in this sense does not have an exact manifestation.

Units of action (units) represent the basic movements performed by individual muscles or a group of muscles.

Action descriptor (DD) – unitary movements performed by muscle groups (for example, pushing the lower jaw forward). The muscular basis for these movements was not specified. And they do not manifest themselves in behavior as precisely as the ED is manifested in it.

For example, a slide can be used for two types of smiles:

1. Deliberate insincere smile: only the main zygomatic muscle contracts
2. Involuntary sincere smile: contraction of the main zygomatic muscle and the lower part of the circular eye muscle.

The experts also managed to develop an ESCLD (Emotional Coding System for Facial Movements), which considers emotions that are associated with facial movements (Table 1.2).

Table 1.2 - Facial movement coding system according to P. Ekman

Emotion	Unit of action
Joy	6+12
Sadness	1+4+15
Surprise	1+2+5B+26
Fear	1+2+4+5
Anger	4 +5 +7 +23
Disgust	9+16+15
Contempt	R12A+R14A

Also, in the matter of defining emotions and pattern recognition, attention should be paid to the task of intelligent video analysis. In such systems, the manifestation of emotions is divided into classes of external signs.

The "emotional state" is compared with the "level of tension" of a person. Emotional tension is tension at the level of the psyche caused by prolonged overload of the emotional sphere without changing the intensity of manifestation.

The task of automatic emotion recognition is attracting more and more attention, and, as a result, more and more different methods are being used to solve it [10].

Currently, the following classes of methods can be distinguished that use profiling elements in the development of a recognition system:

- holistic methods (that is, processing the entire face image);
- local methods (i.e. processing images of face elements);
- methods that calculate the shape of objects;
- methods that calculate the dynamics of objects.

Artificial neural networks (ANNs) and support vector machines (SVMs)

GA – genetic algorithms; LVQ – vector quantization (Kohonen network); AAM – representation model; ASM – shape model; EBGM – elastic graph; PCA – principal component method; HoG stands for histogram of directional gradients; LDA stands for linear discriminant analysis; and LBP stands for local binary patterns [11]. These classes are listed in table 1.3:

Table 1.3 – Classification of emotion recognition methods

Methods	Holistic methods	Local methods
Methods that calculate the shape	Classifiers: Artificial Neural network, Random Forest, Adaboost, Wavelets and Gabor filters Hough Transform 2D Face Models: AAM ASM EBGM	Classifiers: Artificial Neural network, Adaboost Bayesian Classifier Geometric models of the face of the Eigenvector: PCA Local histograms: HoG LBP
Methods for calculating dynamics	Optical flow Dynamic models	3D Dynamic Models Statistical Models: HMM DBN

The merging of the survey script and the theory of microexpressions allowed profiling to move to a different level of development and gradually begin to integrate into economic structures, in particular, into the banking sector, video surveillance security systems.

Thus, profiling is an up-to-date method of personality diagnosis, which significantly increases the level of security for law-abiding citizens.

1.5 Artificial intelligent emotion recognition systems

Using the theory of profiling and basic emotions, as well as the modification of emotions, it is possible to describe a mathematical model of a particular emotion. Any facial expression is analyzed and translated into a language understandable to the intellectual system. The main and related stages of creating a mathematical model are:

1. Identification of faces in the frame;

2. Understanding by the artificial system that in the presence of an unnatural face rotation or insufficient illumination, the person in the frame is the same person.

3. Emphasizing the distinctive features of the face—such as the elongated face, the position of the lips, the cut and shape of the eyes—that enable one to identify a person from another.

All of this is done instantaneously and automatically by the human brain. In actuality, humans are quite skilled at identifying faces and ultimately recognize faces in commonplace items.

Because of their limited capacity for high levels of generalization, computers must be taught each step of the procedure independently.

There are many viewpoints on defining the number of phases in the creation and design of intelligent systems. This relies on a number of variables, including the region of usage, the kind of functions the future intelligent system will have, the accessibility of cutting-edge technologies, and many more.

Let's look at some of the existing software systems for automatic emotion recognition.

To analyze the programs, three criteria were selected in the summary table: system capabilities, advantages and disadvantages (Table 1.4). The "System Capabilities" column also provides the method that was used in the implementation of the system.

Table 1.4– Emotion recognition software

Emotion recognition system	Information about the system	System features	Dignities	Disadvantages
Face Reader	The developer company: Nold us Information Technology (Netherlands), face videography is used	Emotion recognition, determination of ethnicity, use of the Active Template method, creation of an artificial face model.	Recognition with 89% accuracy; emotions can be detected frame by frame or in full (video), full visualization (histograms, diagrams of emotions)	Does not recognize children under 5 years old; inaccurate definition of emotions in a person with glasses; different skin color is perceived differently by the system; a turned face is not detected.
EMotion Software GladOrS ad	Developer company: Visu al Recognition (Netherlands)	The system creates a 3D model of the face with the identification of 12 key areas, such as the corners of the eye and the corners of the mouth.	The system creates a 3D model of the face with the identification of 12 key areas, such as the corners of the eye and the corners of the mouth.	The details of the implementation algorithm are unknown (no flaws have been identified)
FacE Analysis System	MMER-Systems developer Company (Germany)	The imposition of a certain deformable mask on the face, which allows you to calculate the necessary parameters in real time	Recognizes six basic emotions, determines gender, age, ethnicity; identifies a person if the photo was previously uploaded to the database; has additional modules	Incomplete coverage of uploaded data, because you can work with a webcam; inaccurate results on uploading data.
FaceSecurity	Developer company: Cognite c (Germany) Current version: 4.6	The development is designed to process unique databases and databases of any category of people,	Cluster configuration; sorting of the list of images of people; search selections are available	Poorly lit faces are not recognized; only frontally positioned faces are recognized

The process of building artificial intelligent systems is conditionally divided into five stages (Figure 1.7):

Figure 1.7 - The stages of designing an AIS

1. Determining what jobs are defined as well as what qualities they possess. There is a limited user base for the system, and a technical specification is currently being established.

2. Emphasizing the key ideas in the field, which represent the expertise of several specialists. Formal methods of presenting information and methods for solving problems are defined by the knowledge engineer [12].

The ideas that guide the selection of a characteristic scheme for encapsulating the subject-matter expertise of an expert are recognized and developed. A human expert, books, technology descriptions, instructions, papers, techniques for brainstorming, and automatic database filling are the primary sources of information about the subject matter. The Internet is a valuable resource for knowledge acquisition, encompassing both conventional methods of searching for essential information and expertise, as well as clever software robots.

3. Choosing the formalism for knowledge representation and defining the decision-making process. The following step, which is the direct building of the system's knowledge base, is implemented using the knowledge representation structure that has been constructed.

1. Selecting or creating a language for knowledge representation. The knowledge engineer enters the rules into the database once they have been developed and presented in the chosen presentation language.

2. Working through certain verification tasks to test the system.

There is no tight regulation of the processes involved in developing intelligent systems. Drawing a meaningful and chronological division between some of them is challenging. They do, to a certain extent [13], roughly capture the process of building intelligent systems.

Neural network training is a prerequisite for developing an artificially intelligent emotion recognition system utilizing a neural network. You must choose the values of the interneuronal connections' parameters in order to build a neural network for a particular purpose.

This process requires a large amount of data to output sets of facial characteristics and high computer performance.

Conclusions on the first chapter:

1) Artificial intelligent systems are classified by types, tasks to be solved and methods of use. A wide range of AIS allows you to choose the most effective system and the most accurate method for solving the problem.

2) The process of developing an AIS does not have strict regulations. Conditionally, design can be divided into five stages: identification of task characteristics; finding concepts for representing knowledge; designing a structure for representing knowledge; forming rules for representing knowledge; evaluating rules for representing knowledge.

3) A mathematical model based on the concept of biological neural networks, along with its hardware and software implementation, is referred to as an artificial neural network. When attempting to mimic the functions taking place in the human brain, this idea emerged.

4) Modeling the environment in which an artificial neural network is embedded leads to the difficult process of training, which involves adjusting free parameters. One of the many issues with neural networks is determining which model is best for creating a specific system that would enable neural networks to be used in practice.

The use of profiling is advisable to use face recognition in images or video data stream in security services, business, banking and other areas when solving problems.

Chapter 2

Information Models of Neural Networks and Emotion Recognition

2.1 Unidirectional Multi-layer Networks

Formally speaking, a multilayer neural network is just an assembly of basic processing units known as neurons. Layers of neurons are joined by unidirectional connections known as synapses.

A network typically consists of one or more hidden layers of computational neurons, one output layer of neurons, and several sensory components, or input nodes, that compose the input layer.

Unidirectional multilayer neural networks are the most common, as they have a simple mathematical description. Since the late 70s, multilayer networks have been developed thanks to convenient learning algorithms.

Direct propagation neural networks have an input signal that is transmitted from layer to layer (from one neuron to another). It is precisely such networks that are called a multilayer perceptron, which consists of an input layer, hidden computing layers inside the system and an output layer of neurons.

A multilayer perceptron is a unidirectional sigmoidal type network.

The sigmoidal function is an S-shaped nonlinear function with saturation that increases monotonically everywhere differentiable. It is a highly useful activation function to utilize in a formal neuron. You may boost weak signals with the sigmoid and avoid becoming saturated by powerful ones.

The block diagram of a multilayer perceptron (in this case, a two-layer perceptron) looks like this (Figure 2.1):

Notation:

, l=1,2,...,L are signals coming out of the first layer of neurons;

(m) , m=1,2,... - the indexes in parentheses at the top indicate the number of the neuron layer;

1, 2, ... , – the input signals that form the input c1, w_2, ... , w_n are the first hidden layer;the fifth layer;

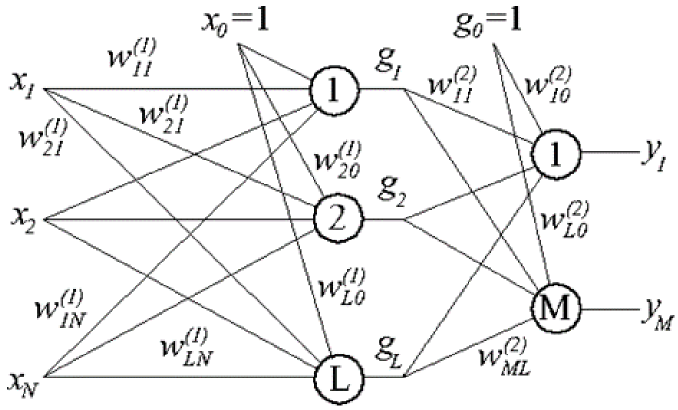

Figure 2.1 - Diagram of a multilayer perceptron

1, 2, ... , – the output signals that form the output layer. The encoding is "N-L-M" for the convenience of designating the network structure .

Formulas for calculating the output signals of the layers of neurons:

$$g_l = f\left(\sum_{j=0}^{N} w_{lj}^{(1)} . x_j\right), l = 1, 2, \ldots.., L \tag{2.1}$$

$$y_i = f\left(\sum_{i=0}^{L} w_{il}^{(2)} \bullet g_l\right) = f\left(\sum_{i=0}^{L} w_{il}^{(2)} \bullet f\left(\sum_{j=0}^{N} w_{lj}^{(1)} \bullet x_j\right)\right), i = 1, 2, \ldots M \tag{2.2}$$

The objective function for a single training sample <X,D> looks like this:

$$E(W) = \frac{1}{2}\sum_{i=1}^{M} (y_i - d_i)^2 \tag{2.3}$$

An objective function for a set of training pairs $< X^k,\ D^k >$, k=1,2,......,p, it has the form of a sum for all pairs:

$$E(W) = \frac{1}{2}\sum_{k=1}^{p}\sum_{i=1}^{M} (y_i - d_i)^2 \tag{2.4}$$

The purpose of training a multilayer perceptron: at the input it is necessary to select such values of the network weights $w^{(1}{}_{lj}{}^{)}$ and $w^{(2}{}_{lj}{}^{)}$, so that the output vector YK and the vector of expected values D^k coincide as much as possible.

2.1.1 Training of a Multilayer Perceptron by the Method of Error Back Propagation

One of the perceptron learning methods is the error back propagation algorithm. It is assumed that the vector will have two passes through the network layers – one forward and the second reverse. In the first case, the vector starts moving from the input layer, then moves along the network from layer to layer. At the same time, a set of output signals is generated inside the network itself.

During the return pass, the synaptic weights are adjusted in such a way that the actual output is subtracted from the desired network output, after which an error signal is received. Further, the signal in the direction opposite to the synaptic connections propagates through the network. That is why this method is called error back propagation.

Let's consider the essence of this method using the example of a two-layer perceptron from Figure 8. The case of a single training sample will help to simplify the example as much as possible, the objective function will look like according to formula (2.1.3). To clarify the coefficients, the formula of the gradient method is necessary:

$$w^{k+1} = w^k - \eta \bullet gradE(w^k) \tag{2.5}$$

In the implementation of this formula, the main difficulty is the calculation of the gradient components of the objective function, formula (2.1.6):

$$E\left(W\right) = \frac{1}{2}\sum_{i=1}^{M}(f(\sum_{l=0}^{L} w_{il}^{(2)} \bullet g_l) - \ d_l)^2 = \frac{1}{2}\sum_{i=1}^{M}(f(\sum_{l=0}^{L} w_{il}^{(1)} \bullet x_j) - d_l)^2 \tag{2.6}$$

Derivatives of the target function by the weights of the output layer neurons:

$$\frac{\partial E\left(W\right)}{\partial w_{il}^{(2)}} = (y_i - d_i) \bullet \frac{\partial f\left(u_i^{(2)}\right)}{\partial u_i^{(2)}} \bullet g_l, \quad u_i^{(2)} = \sum_{l=0}^{L} w_{il}^{(2)} \bullet g_l \tag{2.7}$$

Let $\delta_i^{(2.1.2)} = (y_i - d_i) \bullet \frac{\partial f\left(u_i^{(2)}\right)}{\partial u_i^{(2)}}$, then the component of the gradient vector has the form:

$$\frac{\partial E\left(w\right)}{\partial w_L^{(2)}} = \delta_i^{(2)} \bullet g_l \tag{2.8}$$

Components for the weights of the penultimate layer:

$$\frac{\partial E\left(w\right)}{\partial w_L^{(1)}} = \delta_i^{(1)} \bullet x_j \tag{2.9}$$

The analysis of the formulas allows us to formulate a rule for calculating the gradient vector E(W) for derivatives of the objective function for weights in each layer:

$$\frac{\partial E\left(w\right)}{\partial w_L^{(k)}} = \delta_i^{(k)} \bullet x_j \tag{2.10}$$

X_j - the signal at the entrance

$\delta_i^{(k)}$ – the learning error.

Due to the error transfer *(y_i-d_i)* from the network output to the previous layers?

The method is called reverse error propagation.

The model of a multi-layered sigmoidal network and the method of its training, despite its relative simplicity and convenience, has a number of difficulties that slow down the learning process, or do not allow the network to learn at all. For example, the values of the weights as a result of the correction can become very large values.

This situation will lead to the fact that most of the neurons will output huge values, but the derivative in these values will be very small.

The process may stop because the error sent back in the learning process is proportional to this derivative.

In addition, you need to choose a step size so that it is finite. In this matter, you can rely only on experience. This algorithm has a proof of convergence. The step size will directly affect convergence: either it will be too slow if the step size is very small, or it will undergo constant instability due to too large a step size.

Another difficulty is the temporary instability. The network can recognize letters, but if, after learning one, it forgets the other, then learning will not make sense. The learning process should be conducted in such a way that the network is trained on the entire set, without losing the information that has already been learned.

2.2 Recurrent Neural Network

A recurrent neural network is the most complex type of neural network that has feedback between elements: from more remote to less remote. It is feedback that allows the network to remember and learn. However, the potential of such a network is poorly understood due to the complexity of the analysis .

A recurrent artificial neural network is a network obtained from a unidirectional multilayer perceptron by introducing feedback loops with a delay from the output to the network inputs [46].

Let's consider one of the varieties of such networks - this is the recurrent Elman network, consisting of two layers. The block diagram of this network is shown in Figure 9.

It is clearly seen that the feedbacks come from the inner layers, and not from the outputs of the neural network. It is this structure that allows you to accumulate and remember information.

The signal enters the input and passes into the hidden layer. After conversion by the hidden layer, the signal will go to the output, and its copy will be delayed. Then the next signal is sent to the network, and at the same time a copy of the previous signal is received.

Here , = 1,2, ... , − output signals for the first layer; Standard view of the objective function:

$$E^k(w) = \frac{1}{2}\sum_{i=1}^{m}\left(y_i^k - d_i^k\right)^2 = \frac{1}{2}\sum_{i=1}^{m}\left(e_i^k\right)^2 \tag{2.11}$$

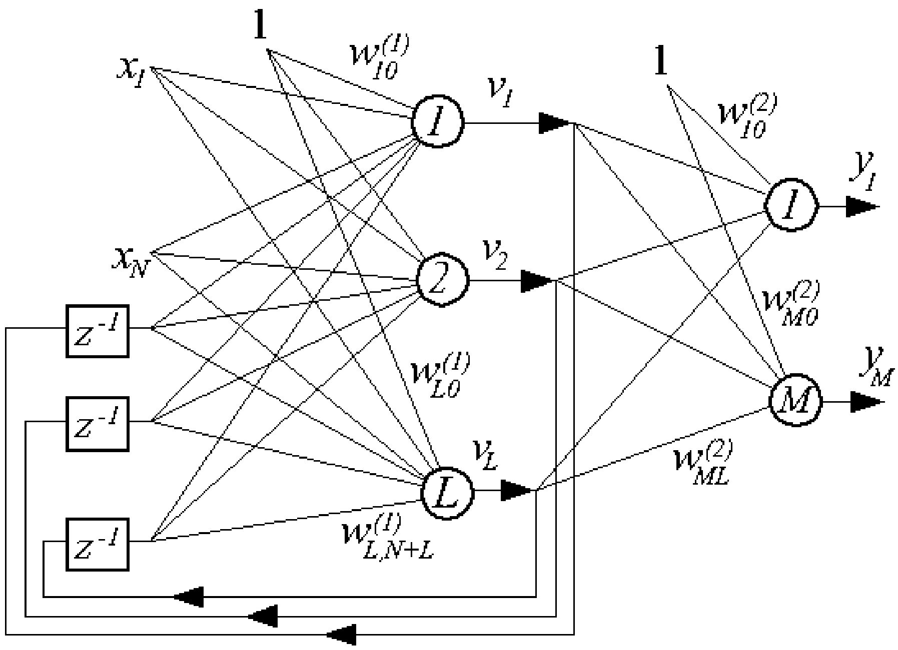

Figure 2.2 – Recurrent neural network

Elman's formula looks like this:

$$\frac{\partial E^k(w)}{\partial w_{nn}^{(2)}} = e_n^k \cdot \frac{\sigma f2^{(}g_{n)}^k}{\partial g_R^k} \bullet \nu_m^k \tag{2.12}$$

The training of the recurrent Elman network can be set using the following algorithm:

1) Determine the initial values for the weights, assign k = 1.
2) Determine all active network signals for the current moment k.
3) Calculate the values for all weights of the neurons of the first layer.
4) Calculate all the components of the vector of the objective function.
5) Adjust the weights of the neurons for all layers.
6) Increase k by 1, go to point 2.

Elman networks are mainly used in control systems for moving objects. They do not provide high accuracy of the solution, since the presence of feedback in the hidden layer does not allow us to accurately calculate the gradient of the functional. However, such networks are capable of processing sequences of images, taking into account the connection between the elements of the sequence; they have a high generalization ability, which is much higher than that of a multilayer perceptron. Recurrent networks are used in audio and speech processing tasks, electronic circuit analysis, computer vision and signal processing tasks.

2.3 Kohonen Networks

A separate class of neural networks, called Kohonen networks, is used to solve various classification problems. The main task of classification is to divide objects into classes, according to any criteria. The basis of this separation is the vector of object parameters. Such networks are also called self-organizing networks based on competition [47]. Neural networks do not require a "teacher", but operate on the principle of "winner takes all". From a mathematical point of view, this means that there is a two-layer network in which each neuron has all the components of the input vector X of dimension N (Figure 2.3). At the output, the largest signal will become a single one, while the rest will turn to zero.

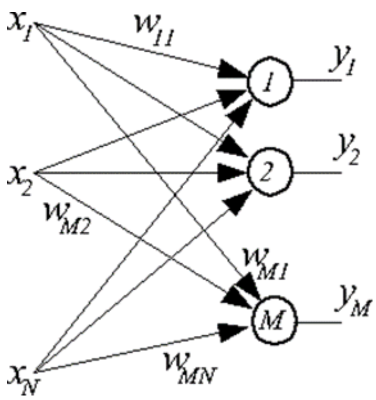

Figure 2.3 – Kohonen Network [

The network is based on WTA (Winner Takes All) type neurons[48], which look like this (Figure 2.4):

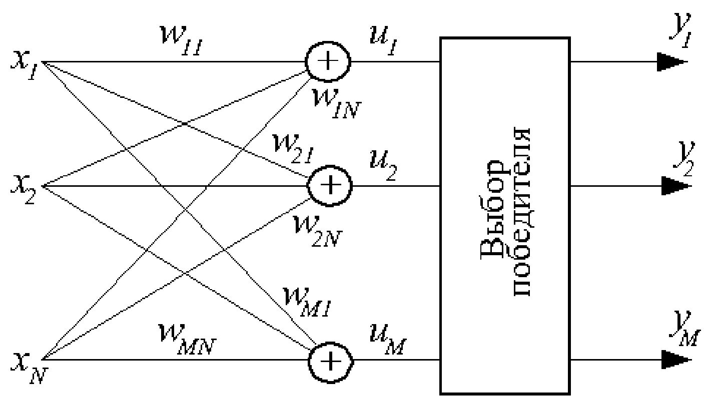

Figure 2.4 – The Winner-takes-all neural network [48]

The input signals for all competing neurons are the same,

the output signals are calculated according to the formula:

After comparing the results, the winning neuron is selected (the highest value). At the output, the signal gets a value of 1, in case of victory, all other signals of neurons get a value of 0.

Let's go back to Figure 10. The weights of the input links form a vector

$$[w_i1, w_i2, ..., w_iN]^{(2.3.1)}$$

. Normalization of the values of the input vector is very often required, because in addition to the connections presented in the diagram, there are connections that indicate the degree of "neighborhood" between neurons.

The learning process looks like this:

$$X_k \tag{2.13}$$

the training vector is fed to the network input;

$$d(X^k, w_i)$$

$$w_i^{k+1} = w_i^k + \mu_i^k \left(x^k - w_i^k\right) \tag{2.14}$$

μ_i^k- the learning rate. With a passion for the distance from the i-neuron to the winner. The coefficient value will decrease. Weights of neurons outside the neighborhood S_w^k it will not change. The size of the neighborhood and the learning rate decrease over time.

2.3.1 Kohonen's Algorithm

In neural networks proposed by T. Kohonen in 1982, neurons form a one-dimensional chain, with each of the neurons having neighbors on the left and right. In a more complex case, neurons can form a so-called grid, where each neuron will have one neighbor on the right, left, top and bottom. Another case is when each neuron can have six neighbors on the plane (like a clock face 2,4,6,8,10,12), that is, the grid will be hexagonal [49].

According to the Kohonen algorithm, the correction of weights is adjusted and expressed by the formula:

$$w_i^{k+1} = w_i^k + \mu_i^k \cdot G^k(i,\ X^k) \left(x^k - w_i^k\right) \tag{2.15}$$

Where $G^k(i,\ X^k)$ – the "neighborhood" function is defined by the Gauss formula:

$$G^k\left(i,\ X^k\right) = \exp(-\frac{{}^2\left(,X^k\right)}{2(\delta^k)^2}\ ) \tag{2.16}$$

Where $\mathrm{d}^2\left(,X^k\right)$ – the distance between the winners and I am a neuron in the k learning cycle . Learning rate μ_i^k and δ^k – Gauss width coefficients decrease during the learning process.

As a result of training, neighboring neurons become representatives of the training data. The advantage of Kohonen networks lies in the clarity of data representation, through one-dimensional or two-dimensional visualization. Kohonen networks are also commonly referred to as Kohonen maps.

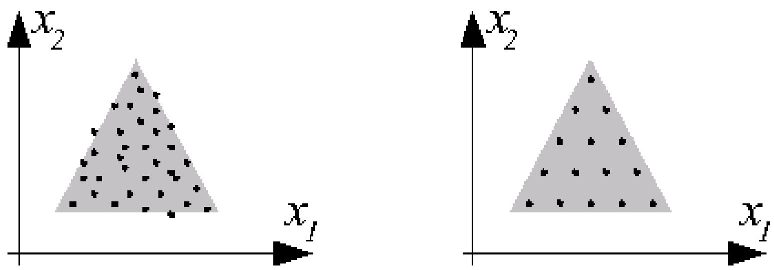

Figure 2.5 – Self-organization of a neural network based on competition

Figure 12 shows an example of training a network with self-organization based on competition. A network with 15 neurons and a two-component input vector is used X=[1, 2]. The figure on the left shows the data in the training sample, on the right the distribution

of neurons in the trained network. The problem is reduced to such a division of space into regions, in which the root-mean-square error is minimized [50].

Kohonen neural networks are two–layer networks consisting of an input layer of neurons and a layer of active neurons (Kohonen layer), which can be one-dimensional, three-dimensional and two-dimensional [51].

Determining the weights of neurons for the Kohonen layer is based on the use of self-learning or clustering algorithms.

Like any model, the Kohonen network also has a number of difficulties in implementation:

1) Kohonen networks work only with numeric data.

2) Initially, you need to set the number of clusters yourself if the clustering method is used.

3) It is necessary to minimize the size of the network.

4) The optimal solution may not always be found. Radial neural networks

The radial basis function network is a neural network of direct signal propagation that contains an intermediate (hidden) layer

of symmetric neurons. Such a neuron transforms the distance from a given input vector to its corresponding "center" according to some nonlinear law [52].

Radial neural networks are built using radial neurons. In the vicinity of its center, the activation function of these neurons has non-zero values. This is called a local approximation.

Approximation is a scientific method consisting in replacing some objects with others close to the initial ones, but simplified in some sense. From a mathematical point of view, multilayer networks of the sigmoidal type approximate the function of several variables ∈ into a set of output variables Y∈ [53].

The sigmoid function plays the role of the activation function of neurons. Since it has a non-zero value over the entire range of input data, all its neurons are involved in converting input data into output within the network. The radial network has two layers: the first layer consists of radial neurons, and the second (output) consists of one neuron or several linear ones. Figure 2.6 shows a diagram of a radial network with one output neuron:

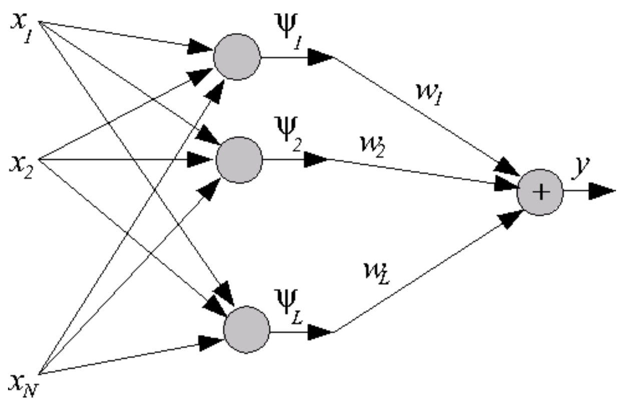

Figure 2.6 – Block diagram of a radial neural network

$(\| - \|)$- the basic function or activation function of the I-th radial neuron. The approximation is performed by the network according to the expression, to simplify $_0 = 0$:

$$y = \sum_{i=1}^{L} w_i \bullet \Psi_i(\|x - C_2\|) \tag{2.17}$$

The task of learning: it is necessary to have such values L, C_i w_i, which deliver the minimum of the target function:

$$E = \frac{1}{2} \cdot \sum_{k=1}^{\mathbb{P}} \left(\sum_{i=1}^{L} w_i \cdot \Psi \left(\left\| X^k - C_i \right\| \right) - {}^k \right)^2 \tag{2.18}$$

When the quantity of training samples, p, is given [54].

Let G be the Green matrix. If the parameters of the radial functions are known, then the learning task will be similar to solving a system of linear algebraic equations G$\bullet$ = ,

where W=[1, 2, ... ,]- vector of weights;

D= [^{1}, ^{2}, ... ,]- the vector of expected values of the network output signal is most often used as a radial function of the Gaussian function:

$$\Psi\left(x\right) = \Psi_{\cdots}^{\left(\|x - Gi\|_2\right)} = \exp\left(-\frac{\|x - C_i\|_2^2}{2 \cdot \sigma^2}\right) \tag{2.19}$$

σ_i- the parameter that sets the width of the function

The learning process of the radial network includes two stages:

1. For each radial neuron, it is necessary to select the parameters of the function Ψ;

2. Weights must be selected for the output layer of neurons.

The second stage is somewhat simpler: if you use the Green matrix G, then the computational costs are to calculate the inverse matrix; if you follow the Gauss formula, then the solution process again splits into two subtasks:

a) determining the centers of .

This refers to the definition of a value for which the main requirements are completeness of coverage of the definition area and uniformity of data distribution.

b) calculation of the parameter

After all the centers of the radial function have been determined, the parameters that determine the magnitude of the coverage area are selected. To simplify the task as much as possible, you need to take the Euclidean distance from the center to the nearest neuron as the value .

Radial neural networks facilitate the selection of initial conditions for the learning process and ensure the rapid achievement of an optimal solution.

The multilayer structure of sigmoidal networks with a single hidden layer is comparable to the radial networks' resultant architecture. In it, fundamental radial functions—which are shaped differently from sigmoidal functions—play the role of hidden neurons. These kinds of networks differ from one another fundamentally, even if certain commonalities have been observed.

Networks of this kind have a weaker generalization ability compared to sigmoid networks. This disadvantage is compensated by dividing the training classes into subclasses. In this case, each subclass will have its own ”center” around which the radial basis function is implemented. In addition, the radial network learning algorithm is unique.

If there is only one hidden layer and the neuron is closely connected to the training data area, the point at the initial stage of training turns out to be closer to the optimal solution.

Table 5 presents a comparative analysis of the second chapter on selected information models of neural networks.

The table lists the main information models of neural networks (horizontally), vertically lists the main tasks that can be solved using artificial neural networks.

Table 2.1 – Comparative analysis of neural network models

INS / Tasks	Unidirectional multi-layer networks	Recurrent neural networks (Elman network)	Kohonen Networks	Radial neural networks
Associative memory	+	+	-	+
Information compression	+	+	-	-
Forecasting	+	+		+
Optimization	-	+	+	
Classification	+	-	+	+
Clustering	-	-	+	+

Associative memory is a special kind of machine memory used in very fast search applications.

Information compression is an algorithmic compression of information to reduce the occupied volume. It is used for the rational use of data storage and transmission devices.

Forecasting is an assessment of possible development paths based on accumulated experience and current assumptions.

Optimization is finding the best solution to a problem under given conditions and constraints.

Classification is the solution of a problem by dividing objects into classes according to some criteria.

The goal of the task of clustering is to divide a given selection of items into disjoint subsets, or clusters, so that the objects in each cluster are comparable.

2.4 Intensity of Action of Facial Muscles Movement

The following formula is used to calculate the intensity of muscle movement: the SCLD unit number plus the letters A through E are added based on the movement's intensity (minimum to maximal).

Principles:

a) A is just barely discernible;

b) B— minor;

c) C – apparent or strong;

d) D— acute or very apparent;

e) E is represented as fully as possible;

In relation to the hypothetical vertical axis of the face, motor units can be left, right, one-sided, symmetrical, or two-sided.

The list of the main motor units is presented in 2.1:

Table 2.2 – Motor units and motor descriptors

Action unit number	The name in the SKLD	Muscle Foundations
0	A neutral person	–
1	Raise the inside of the eyebrow	Frontal (medial part)
2	Raise the outer part of the eyebrow	Frontal (medial part)
4	Lower your eyebrows	the muscle that lowers the upper lip, the muscle that lowers the eyebrow, the muscle that shifts the eyebrows
5	Lift the upper eyelid	the muscle of the eye socket that raises the upper eyelid
6	Raise your cheeks/cheek	circular eye muscle
7	Pull on the eyelid/eyelids	The circular muscle of the eye is a mimic muscle that closes the eye slit and dilates the lacrimal sac
9	Raise the wings of the nose	the muscle that lifts the upper lip and the wing of the nose
10	Lift your upper lip	the muscle that lifts the upper lip
11	Deepen the nasolabial fold	The facial muscle that deepens the nasolabial fold
12	Lift the corners of your lips	large zygomatic muscle

Determining the intensity of movement of facial muscles, as well as knowledge of the main facial expressions on a person's face is an integral part in the development of an intelligent system. The developer needs to describe each encoded facial expression in a geometric and mathematical representation. This step is the most difficult, as it requires processing a large amount of information and a sufficient amount of knowledge in the chosen field.

2.5 Analysis of Face Detection Methods in the Image

The task of emotion recognition is a task of a high technical level and is effectively used in automatic accounting systems for the number of visitors; access control systems at airports, subways, enterprises; automatic accident prevention systems; intelligent human-computer interfaces, as well as to stabilize the face image in order to facilitate emotion recognition.

Algorithms for detecting faces in an image can be divided into four categories:

- empirical method;
- the method of invariant features;
- recognition based on a template implemented by the developer;
- the method of detection by external signs (training systems) [55].

Empirical reasoning is based on top-down knowledge and requires the implementation of an algorithm involving the rules that can be attributed to the image fraction where one can find a human face. A set of the rules is a formalization of the empirical knowledge about the face representation in the image and the signs, which lead a human to the decision if he can see a face or not. My rules include the following ones:

- the central part of the face has uniform level of brightness and color; - the splash of the brightness level of central and upper parts is significant; - the face consists of the two symmetric located parts for the eyes with the level of brightness of the eyes is much different from that of the face, nose and mouth.

The image reduction method, in order to eliminate possible interference, as well as to reduce computational operations, pre-exposes the image to a strong size change (Figure 2.8):

Figure 2.8 – Image reduction method [55]

For the given image, the area of uniform brightness distribution is to be with first determined, i.e., the assumed place of the face, and the presence of sharply different zones of brightness inside is to be checked. It is these zones that can serve as a “face” with a certain probability. The histogram method to find the areas of an image containing a “face” constructs vertical and horizontal histograms.. Using this method, the following search for facial features in “suspicious” areas is possible (Figure 2.9):

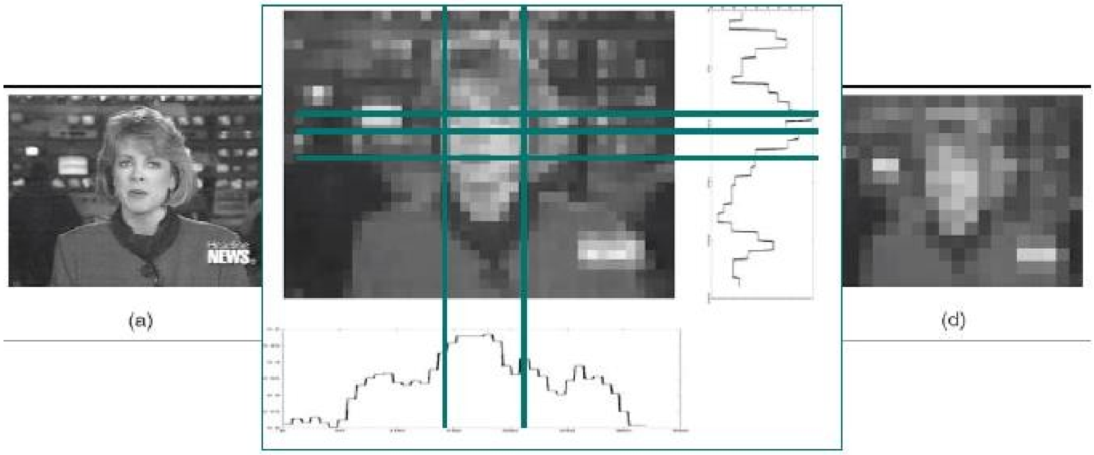

Figure 2.9– The method of constructing histograms

This approach was used in the development of machine learning because it assumed low processor power requirements for image processing.

The presented methods have good indicators for detecting a face in an image with a homogeneous background, they are easily implemented using machine code.

Later, many algorithms were developed. However, it should be noted that these methods are absolutely unsuitable for processing images containing a large number of faces or a complex background. They are also very sensitive to the tilt and turn of the head.

Methods of characteristic invariant features based on bottom-up knowledge form the second family of face recognition methods.

Here, the approach to the problem is indicated in the form of a lack of formalization of the process occurring in the human brain. Proponents of the approach try to identify patterns and properties of the face image implicitly, to find invariant features of the face, regardless of the angle of inclination and position.

The main stages of the algorithms of these methods are:

- detection on the face image: eye, nose, mouth;
- Detection: face borders, shape, brightness, texture, color;
- combining all found invariant features and verifying them.

The method of identifying “complex” faces is based on the search for the correct geometrically arranged facial features. This method uses a Gaussian filter where many different scales and orientations are selected. Then, the search for matching the detected characteristics of the site and their location occurs.

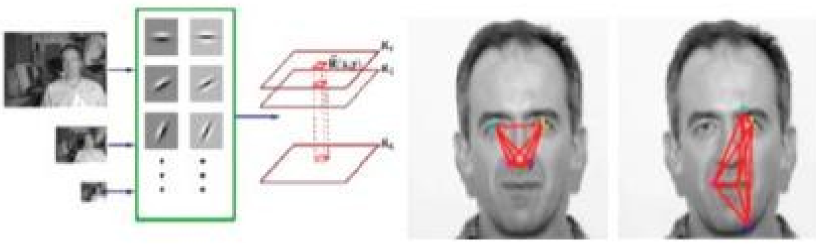

Figure 2.10 – Method of detecting ”complex” faces

The feature grouping method is based on the method of the second derivative of a Gaussian filter, which is selective with respect to which parts of the image should be examined. After that, the edges are grouped by thresholding filter around each zone. Finally, a Bayesian network appraisal is applied to combine the identified features and establish a sample of facial features : (Figure 16):

These methods are able to recognize a face in different positions. But if the face is slightly cluttered with other objects, noise occurs, or the color is too light, the recognition accuracy drops significantly. The saturated background also has a significant impact. The basis of the considered approaches is empiricism, which is both their strong and weak side.

The application of rules of thumb allows you to build a certain model of the face image and reduce the task to performing a certain number of relatively simple checks.

Recognition according to the developer's template involves setting a standard image image by describing individual properties and areas of the face, as well as their possible mutual location [55].

Face detection using a template consists of checking each of the image areas for compliance with a given template. Let's consider two types of templates: non-deformable and deformable.

The templates are pre-programmed, they use correlation to find faces in the image.

The method of decomposing a face using three-dimensional shapes involves using a template in the form of pairs of brightness ratios in two areas. To determine the location of a face, it is necessary to go through the entire image for comparison with a given template using a different scale (Figure 2.11):

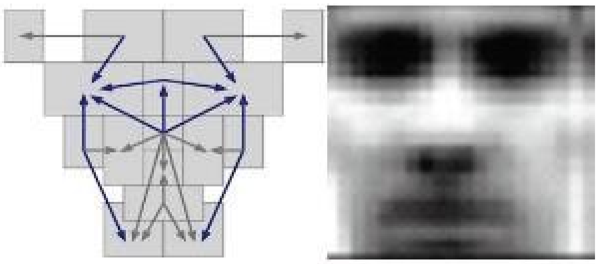

Figure 2.11 – Decomposition of a face into three-dimensional shapes

The control point distribution system is a bending-shape statistical model of incidents. The greatest benefit to this method is that it allows for the distribution of what would normally be a variable objects over a small parameter learning set. Feature classification systems adopt this method.

Pattern recognition is straightforward to implement and works well with pictures having a clear background. The drawback of the approach is that it calibrates this template around the face image. The next method is the detection of a face by external signs (with a training stage and processing of test images). The image system determines the ratio of the vector to one of the classes – face / non-face. The standard search for faces in an image using a mathematical model consists in an absolute search of rectangular fragments of the image. The absolute search scheme has a disadvantage – redundancy and computational capacity of the process, therefore it is necessary to reduce the number of fragments in the image [55].

This is the most common method to solve image recognition problems known as neural networks approach. When solving the problem, we use Support Vectors to reduce the feature space dimensions. At the same time, method support vector — does not lead to a loss of information content of selected training objects and also allows switch on some space base where variance will be oriented along n base axes.

The subspace with which the main axes obtained in this manner is then expanded over ll spaces best instantiated on the training set (Fig. 2.12)

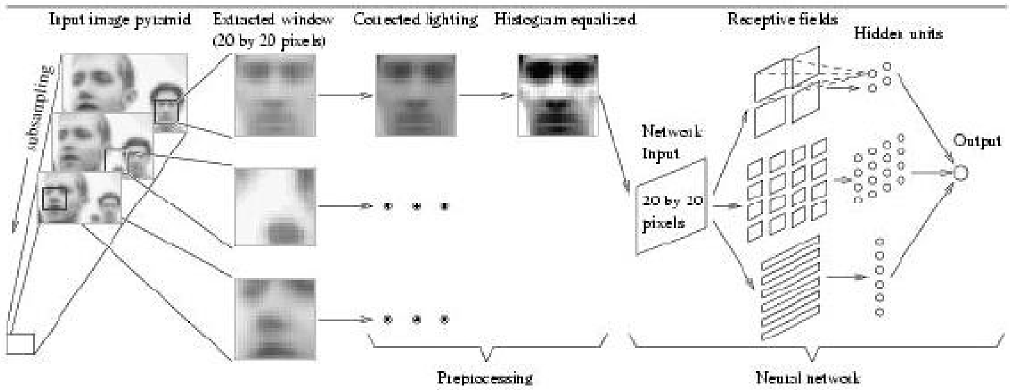

Figure 2.12 – Training an intelligent system using a neural network

These learning algorithms are also used for classification and regression analysis tasks as they are behalf on supervised learning technique or in other words we can say these algorithms perform well when trained data has label and target values. As it is intended for that the support vector machine methodology distributes with classes linearly.

The idea behind training of many classifiers is to reduce the error in classification on the training set (or empirical risk).

The Viola–Jones method is by far the most promising method of pattern recognition due to its high performance and high percentage of face detection. The features used by the algorithm use the summation of pixels from rectangular regions.

Figure 2.13 – Diagram ”Using face detection methods”

The analysis of methods for detecting faces in an image allowed us to identify effective methods for detecting faces in an image, determine the advantages and disadvantages of each method, identify what kind of images can be used and effective in recognition, evaluate each of the methods from an expert point of view.

To date, the Viola-Jones algorithm is the most in demand due to its high speed and high accuracy of operation. Neural networks ranked second in percentage terms. However, it should be repeated that the difficulty of using neural networks lies in their mandatory training.

Conclusions on the second chapter:

1) In the second chapter of the diploma, four models of neural networks are considered:

- unidirectional multi-layer network;
- recurrent neural network (Elman network);
- Kohonen Network;
- radial neural network.

Each of the models has its own set of characteristics, an implementation option and a mathematical description. When choosing a neural model, one should take into account: the structural representation of the network, the selection of input data, the possibility of using a neural network for the selected system, the network's ability to learn, the possibility of finding the optimal solution to the problem.

2) A multilayer sigmoidal neural network based on a multilayer perceptron has a simple mathematical description, but requires competent preparation of the training sequence, formation of a training sample, selection of initial values.

3) A recurrent neural network is a rather complex type of network. It has feedback between the elements, which allows the network to accumulate and store information. Such a network does not always provide high accuracy of the solution, due to the complexity of calculating the gradient of the functional in the hidden layer. If the network is spared from excessive or insufficient data processing, it will also be able to track changes in the readings of the environment in which it was placed.

4) The main task in the implementation of Kohonen networks is to divide objects into classes, according to any criteria. Such networks operate on the principle of "winner gets everything", that is, the output signal of the network with the highest value will become a single one (winner), the rest will turn to zero. Kohonen networks work with numerical data, at the initial stage it is necessary to minimize the size of the network (that is, reduce the volume in the network, in order to facilitate the learning process).

5) Radial neural networks facilitate the selection of initial conditions for the learning process. If there is one hidden layer and the neuron is connected to the corresponding area of the training data space, the learning start point turns out to be much closer to the optimal solution, compared with other neural network models. Radial neural networks have simple learning algorithms. Splitting an object into classes

compensates for the weak generalization ability of radial neural networks.

6) P. Ekman's facial movement coding system is used to describe emotions, For

the introduction of SCLD into an intelligent system, it is necessary to implement a prototype of a facial feature

7) The analysis of face detection methods in the image allowed us to identify common methods that are most effectively used to solve recognition problems. The method using neural networks takes the second place in terms of the effectiveness of its application

8) The analysis of neural network models allowed us to identify two universal models applicable to a wide range of tasks: unidirectional multilayer networks and radial neural networks.

In the practical part of the study, the recognition algorithm will be described and the experimental part on emotion recognition will be performed. A neural network based on a multilayer perceptron was chosen to solve the problem.

Chapter 3

Emotion Recognition Algorithm Using a Neural Network

3.1 Mathematical Formulation of the Recognition Problem

Every person in the world has a range of facial reactions that fall under two categories: geometric (set standard parameters by which signs manifest) and behavioural.

A facial movement coding system is utilised to characterise the parameters of the (voluntary and involuntary) face in a quantitative, qualitative manner. The quantitative parameter in this case is its Strength of the movement from A to E

A video stream is a sequence of frames. Recognition: The goal of recognition is to cluster the faces in images into separate classes. Face recognition We can pose the face recognition problem very simply: assuming you have a know person (the example to be recognized), how can build a function to recognize new images from other examples F(w)=($_1$(), $_2$() ... , ()), the output that defines the class of the image w represented by the feature vector ($_1$(), ... , ()) In this case, the class is one of the six basic human emotions.

The search for a solution is carried out using artificial neural networks.

3.1.1 Obtaining Face Recognition Invariants

An invariant is a property of a certain class (set) of mathematical objects that remains unchanged during transformations of a certain type [59].

Invariant moments are characteristic features that can occur in each image. Most often, the faces on video frames are subjected to various deformations peculiar to human facial expressions. In such conditions, it is necessary to speak about "pseudo-invariants" [60].

The central points of a digital face image are determined by the formula:

$$m_{pq} = \Sigma_x \Sigma y (x - \vec{x})^p (y - \vec{y})^q f(x, y) \tag{3.1}$$

– the central points of the order are not higher than *(p+q)=3, f(x,y)* - brightness function. It is advisable to convert color images to a grayscale view. After preprocessing and normalization, the sample is a matrix of pixels, each of which has a brightness value in the range [0.1].

3.1.2 Choosing an Adequate Metric

To solve the problem of determining whether w belongs to a class let's apply the convolution in the form of a Euclidean measure:

$$(w, \Omega_p) = \sqrt{\Sigma_{L=1}^{7} \cdot \left(M_2^w - M_L^{\Omega} P\right)}^2 \tag{3.2}$$

Where $M_L^{\Omega} P = \left(\sum_{l=1}^{m} M_{i,l}^{\Omega_p} \right) / m_i$ - the average of the invariant moments of all images included in the class Ω_p

The metric can be used to measure the distances between the image w (represented by the feature vector) and the classΩ_p

$$d_{E-M}\left(w\Omega_P\right) = \sqrt{\left(x - \overline{y}\right)^T \left(s_p + E\right)^{-1} \left(x - \overline{y}\right)} \tag{3.3}$$

Where s_p – , $(\overline{y})$ – the center of the class.

3.2 Methods of Facial Motor Activity Recognition

The solution of the emotion recognition problem relates to the classification problem, i.e. the neural network must attribute the received data set to emotions corresponding to a given set of parameters. Let's consider the mathematical description of the recognition problem:

Let's give a set of M images of faces (emotion, for example, surprise) w1, ...wn, each of which has a vector of feature values (facial features) Xi = (xil, ... xim), i=1, ...m, x, j=1, n, where n is the number of features. Feature vectors are assigned by experts to certain classes.

The entire sample is divided into two disjoint subsets: training and test. After training an artificial neural network, the quality of its training on a test set. An artificial neural network of direct propagation with a sigmoidal function of the activity of neurons of the hidden layer and a linear activation function of the output layer is proposed.

The neural network should be configured so that when the feature vector is applied to the input assigned to the class, the network outputs the value "1" at the output with the number I, and "0" at all others. This is achieved by setting up a network of error back propagation methods.

Taking into account the peculiarities of the formation of classes for face recognition, a probabilistic neural network is of interest. In it, samples are classified based on estimates of their proximity to classes, taking into account the features of the probabilistic distribution of feature values. For each class, based on the training data, the density function of the attribute distribution is determined, which is characterized by mathematical expectation and variance.

3.2.1 Methods of facial motor activity recognition

1) To solve the recognition problem, an information model of a neural network based on a multilayer perceptron with reverse error propagation was selected.

2) To create a prototype of emotions (metrics), P. Ekman's Facial movement coding System was chosen.

3) 6 basic emotions were selected for the study: joy, sadness, anger, disgust, surprise, fear. This set allows you to cover the maximum number of real facial expressions.

Each emotion has a prototype of expression, for example, the formula of surprise:

1+2+5B+26 (**??**)

In this formula, motor units are expressed in numbers, each of the numbers characterizes a facial expression that involves a small part of the facial muscles.

The Latin "B" stands for traffic intensity. The neural network model is trained on a large number of images or a continuous video stream in real time.

It should be noted that some of the intensity codes of action are difficult to register, for example, codes A and B have a slight difference in the appearance of facial expressions (Figures 3.1, 3.2):

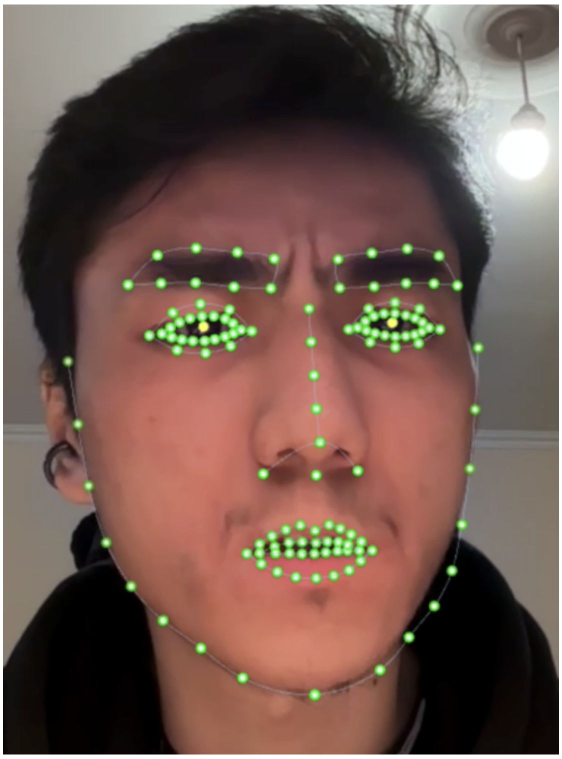

Figure 3.1: Surprise with intensity A

1. The motor units of the face (hereinafter referred to as DE) can be conditionally divided into three groups:

a) group 1: static DE – it is possible to determine by photography.

b) Group 2: dynamic DE – it is necessary to track the change in the position of key points on the face or have an average distance value for this DE;

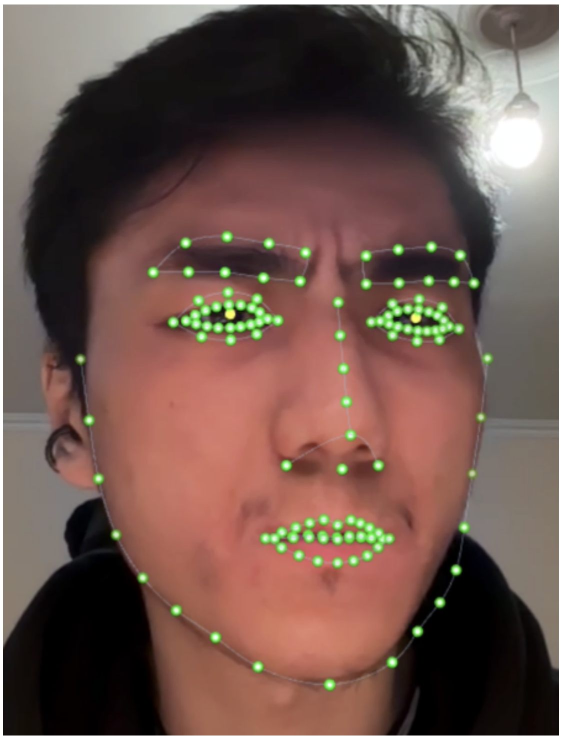

Figure 3.2: Surprise with intensity B

c) group 3: empty ones are actively involved in the manifestation of emotions, but are not registered by search algorithms (dimples on cheeks).

Up to 80 significant facial areas can be identified on a person's face. As a rule, these are the boundaries of the eyes, mouth and eyebrows. The zygomatic muscles are not an important sign of the expression of an emotion.

Each of the emotions has its own dynamics of "experience". Let's consider the components of the groups that are marked in Table 3.1:

Table 3.1 – DE groups for search algorithms

	Static	Dynamics	Empty
Joy	12	6	-
Surprise	1,25	2,5,26	27
Anger	4,25	5,23,24,26	7,10,22
Disgust	15	9,26,10	16,17
Sadness	1,4,15	-	11
Fear	1	2,5,20,38	-

Key metrics system for recognition:

1) Mouth height: upper upper lip – lower lower lip.

2) The height of the open mouth: the bottom of the upper lip is the top of the lower lip.

3) Lip corner down: the corner of the mouth is the top of the lower lip.

4) Lip corner up: the corner of the mouth is the top of the upper lip.

5) Mouth width: the left corner of the mouth is the right corner of the mouth.

6) Chin height: the bottom of the lower lip is the chin.

7.) Eye width: top of the eye – bottom of the eye.

8.) Eyebrow height: the upper center of the eyebrow is the middle of the eyes.

1) Inner corner of the eyebrow: the inner corner of the eyebrow is the inner corner of the eye.

10.)Outer corner of the eyebrow: the outer corner of the eyebrow is the outer corner of the eye.

The metrics numbered 3, 4, 7-10 are asymmetric and are calculated for the left and right halves separately. The normalization of indicators is carried out by the distance between the pupils. It varies minimally from person to person and standardizes the training sample well.

An example of a metric representation is shown in Figure 3.3:

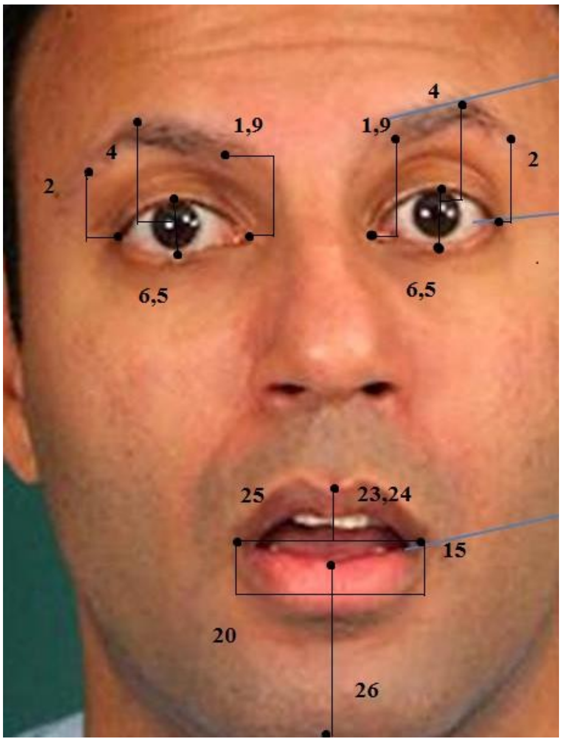

Figure 3.3 Motor units in the metric system

The values of the selected metrics:
1-raise the inside of the eyebrow;
2 – lift the outer one. part of the eyebrow;
4 – lower the eyebrows (the muscle that lowers the upper lip, the muscle that lowers the eyebrow, the muscle that shifts the eyebrows);
5 – lift the upper eyelid;
6 – circular eye muscle;
9- the muscle that lifts the upper lip and the wing of the nose;
15 – the muscle that lowers the corners of the mouth;
20 - facial muscle that pulls the corner of the mouth outward and deepens the nasolabial fold;
23 – the muscle compressing the lips, pulling them forward and closing the mouth;
25 - the muscle lowering the lower lip;
26 –the muscle that lifts the lower jaw and pushes it forward, the chewing muscle.

The decomposition of the image into components, each of which has a certain weight, is shown in Figure 3.3 (surprise emotion):

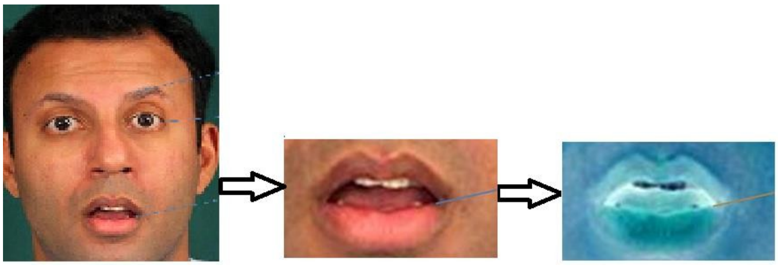

Figure 3.4 — Decomposition of the image into three-dimensional shapes

The decomposition of an image into components is necessary to form input data for a neural network.

Figure 3.5 – Coordinate plane

If we imagine two input vectors of a perceptron on a coordinate plane (Figure 3.5, 3.6), then its job will be to determine on which side of the dividing line lies the vector presented for recognition:

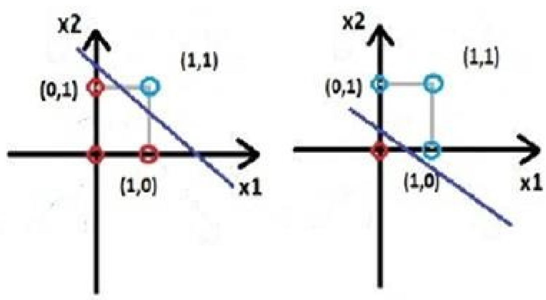

Figure 3.6 – Hyperplanes

As an example, the truth table shows the values of the output signal for the logical functions ”C” and ”OR” depending on the combination of input signals. Correspondence tables for coordinates on hyperplanes:

Table 3.2 – Correspondences to the logical ”C” values

Entrance 1	Entrance 2	Exit
0	0	0
0	1	0
1	0	0
1	1	1

Table 3.3- Correspondence to the values of the logical "OR"

Entrance 1	Entrance 2	Exit
0	0	0
0	1	1
1	0	1
1	1	1

The perceptron can only classify images that are separated by a hyperplane. Let's formulate an algorithm for the back propagation of the error:

1. Initialization of link weights: small random values and maximum standard error;
2. Submit the input vector to the network input;
3. To propagate signals in accordance with direct connections;
4. Error calculation and error of the output layer of neurons;
5. Calculation of the error of the inner layer of neurons;
6. Updating the link weights of each layer

The following is an experimental part of neural network training in the MATLAB program

The neural network was trained in an interactive environment for programming and numerical calculations, as well as visualization of results - MATLAB. Using MATLAB, you can analyze data, develop algorithms, create models and applications [58].

Data for a neural network using a multilayer perceptron:

Number of input neurons: 16 ($4\times$ 4 matrix) Output neurons: 6 (number of emotions); Neurons in the hidden layer: 11.

The number of neurons in the hidden layer is equal to half the sum of the input and output neurons. There is an offset vector for each layer of the network. Let's denote the vector of the hidden layer b1 (includes 11 neurons). The output layer includes 6 neurons - b2.

The result of the sum of the displacement of each neuron is an argument of the neuron activation function. Let's take the number of observations about 3/5 of the maximum possible.

Since 16 binary operations make up the input vector, the number of possible combinations will be 216 = 65536. The training sequence should be approximately 65536 input images. To train a neural network, it is initially necessary to obtain offsets and weights that minimize the learning error.

The algorithm is performed strictly according to a drawn-up plan and a certain set of actions is performed on each cycle:

1) The elements of the training sequence are fed to the network input one by one.

2) The elements of the training sequence are compared with the target (desired) values.

3) The error functional is calculated.

4) The values of the functional and its gradient are used to adjust the weights.

5) Operations are repeated until a certain number of cycles are completed or until the error stops decreasing.

The gradient learning algorithm is the fastest available algorithm in the MATLAB software package, so its use is rational for network training.

To generate the values of the output vector, you need to choose an activation function for which the range of output signals would be defined from 0 to 1. For this case, we choose a logarithmic sigmoidal activation function. This function is also suitable for neurons of the hidden and output layers.

Learning by the method of error back propagation requires the creation of a two-layer neural network of direct signal transmission with sigmoidal activation functions. Let's write it in the command line of the program:
net=newf([01;01;01;01;01;01;01;01;01;01...33[0-1]],[11,6],{'logsig','logsig'},'trainrp');

Explanation:

a) The network has one input vector with 16 elements, for which the acceptable value boundary is the same [0 1];

b) In the first hidden layer, the network has neurons, in the output layer there are 7 neurons;

c) Activation functions used: logsig- calculates the output of the layer by its input, specified for each layer; trainrp – the function of learning and modifying the weights and offsets of the network, in accordance with a given learning algorithm.

Setting the parameters of the learning algorithm:

net.trainparam.goal=1e-5; net.trainparam.show=50; net.trainparam.epochs=5000;
goal - the limit value of the learning criterion, network error;

epochs - the maximum number of training cycles;

show- the information output interval.

inputs=[0...1; 0...1; 0...1; 0...1; 0...1; 0...1; 0...1; 0...1; 0...1; 0...1];
targets=[0...1; 0...0; 0...0; 1...0];

The result: the network has been configured, a training sequence has been formed, and a learning algorithm has been selected.

The learning process of the network: In the MATLAB command line, enter the following expression:*net=train(inputs,targets);*

Starting the neural network learning process (Figure 3.4):

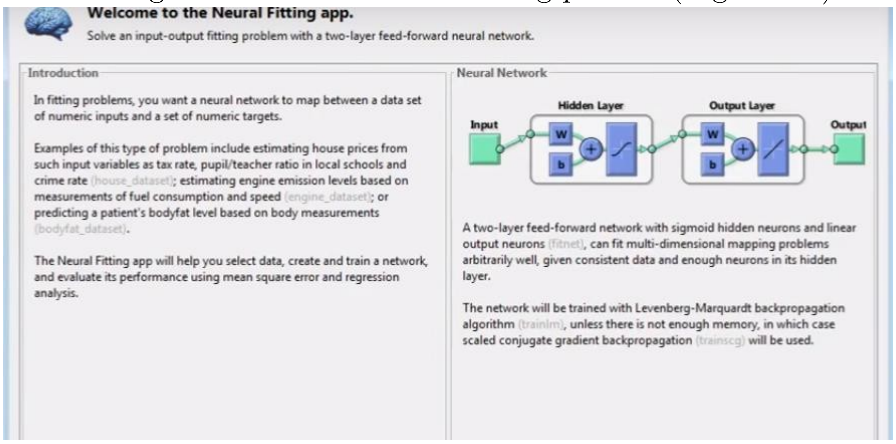

Figure 3.4 – Starting the learning process in MATLAB

Selecting the input and output data file (Figure 3.5):

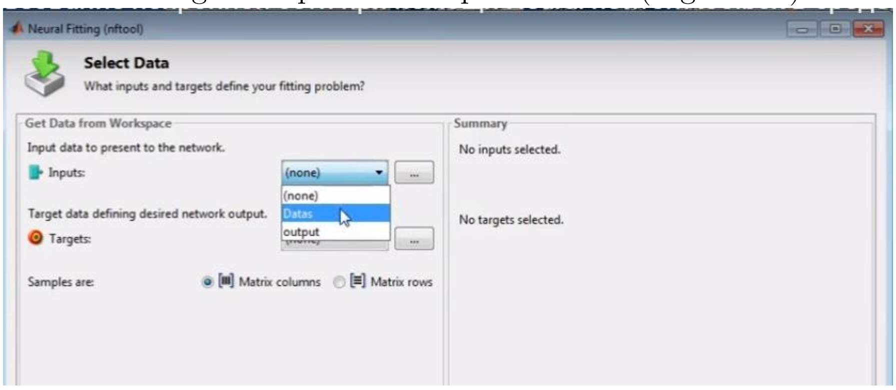

Figure 3.5 – Selecting data from the generated file

Choosing the number of neural layers (Figure 3.6): by the author:

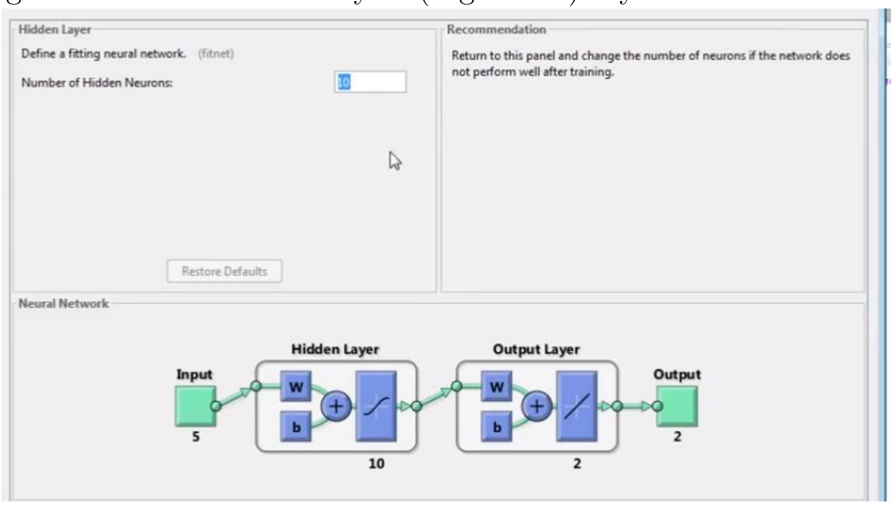

Figure 3.6 – Selection of neural network layers

Choosing a neural network learning algorithm (Figure 3.7):

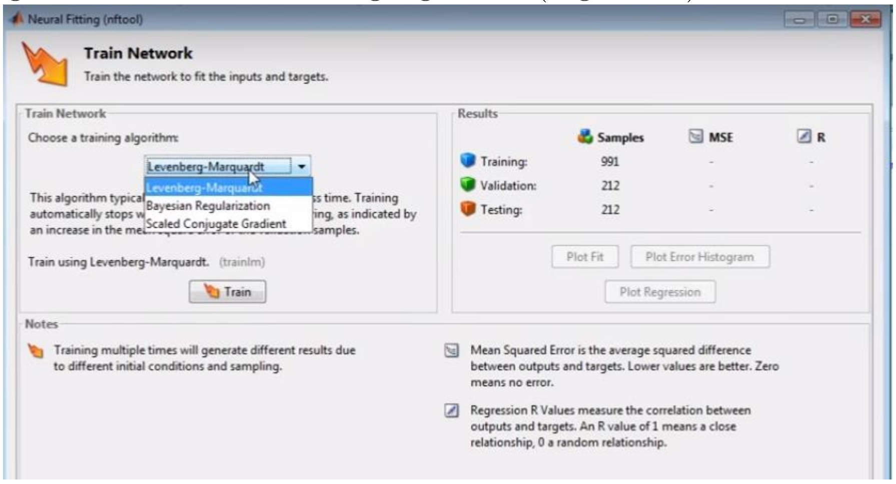

Figure 3.7 – Neural network learning algorithm

The learning process of a neural network (Figure 3.8):

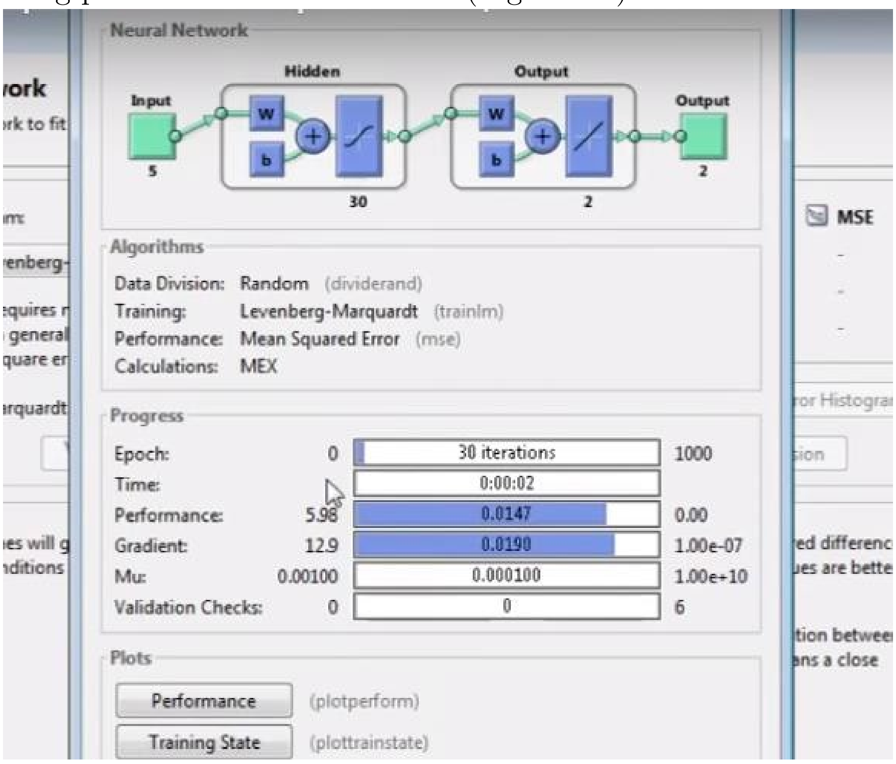

Figure 3.8 –Neural network training

The window with the learning abilities of the network (3.9)

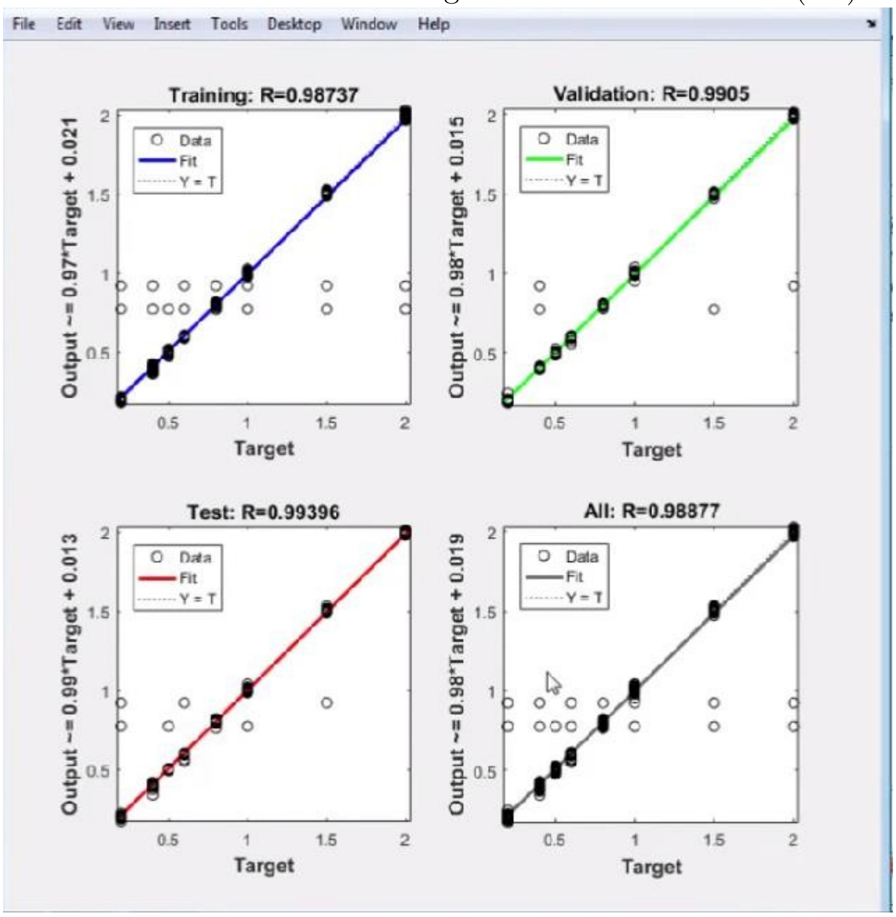

Figure 3.9– Neural network training graphs

During the experiment, an emotion was considered recognized if it had the greatest value at the output of the neural network. As a result of modeling a two-layer neural network, two weight coefficient matrices IW- for the hidden layer, LW- for the output layer and two displacement vectors b1 and b2, respectively, were obtained [15]. For a neural network, it is necessary to build a mathematical model and analyze its functioning (shown in the figures. 3.10, 3.11).

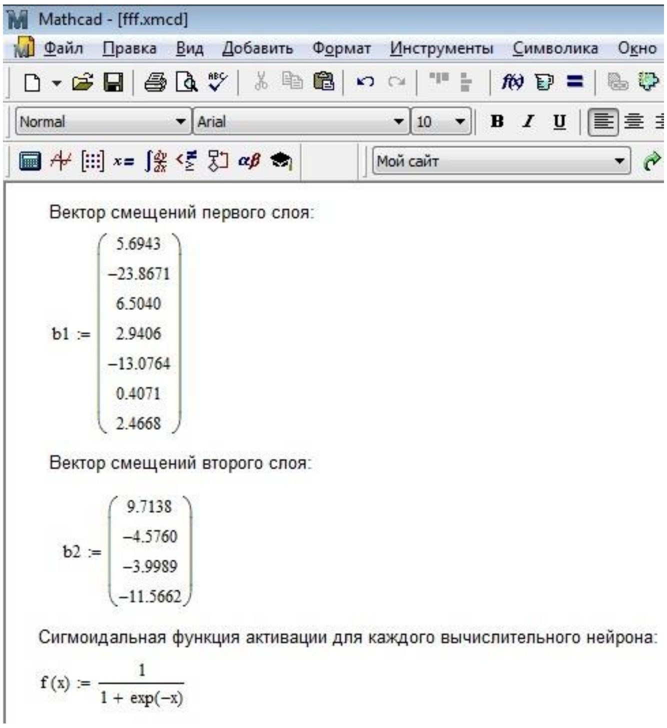

Figure 3.10 – Matrix of coefficients of layers of neurons

2. Расчет вектора аргументов для функций активации нейронов первого слс

$S1 := P^T \cdot IW^T + b1^T = (43.434\ -3.066\ -4.863\ -0.97\ 6.584\ 26.534\ 5.383)$

3. Расчет выходного сигнала A первого (скрытого) слоя нейронов:

$$A := \begin{pmatrix} f\left(\sum S1^{\langle 0\rangle}\right) \\ f\left(\sum S1^{\langle 1\rangle}\right) \\ f\left(\sum S1^{\langle 2\rangle}\right) \\ f\left(\sum S1^{\langle 3\rangle}\right) \\ f\left(\sum S1^{\langle 4\rangle}\right) \\ f\left(\sum S1^{\langle 5\rangle}\right) \\ f\left(\sum S1^{\langle 6\rangle}\right) \end{pmatrix} = \begin{pmatrix} 1 \\ 0.045 \\ 7.67\times 10^{-3} \\ 0.275 \\ 0.999 \\ 1 \\ 0.995 \end{pmatrix}$$

4. Расчет вектора аргументов для функций активации нейронов второго сло

$S2 := A^T \cdot LW^T + b2^T = (-218.432\ -11.037\ 3.738\ -103.985)$

5. Расчет выходного сигнала Y второго (выходного) слоя нейронов:

$$Y := \begin{pmatrix} f\left(\sum S2^{\langle 0\rangle}\right) \\ f\left(\sum S2^{\langle 1\rangle}\right) \\ f\left(\sum S2^{\langle 2\rangle}\right) \\ f\left(\sum S2^{\langle 3\rangle}\right) \end{pmatrix} = \begin{pmatrix} 0 \\ 0.00002 \\ 0.97674 \\ 0 \end{pmatrix}$$

Figure 3.11 – Displacement vectors of the neural network

Result: in the process of modeling a neural network in the MathCAD software environment and applying a vector to the input that corresponds to the emotion "surprise", the network generated the correct output signal. The intensity of the emotion mattered the most. The result of the analysis can be written as follows :

$$\begin{aligned} S1 &= P * IW + b1 \\ A &= f(S1) \\ S2 &= A * LW + b2 \\ Y &= f(S2) \end{aligned} \qquad \text{или} \qquad Y = f(f(P * IW + b1) * LW + b2)$$

Figure 3.12 – A model for determining the emotion of "surprise"

Designations:
P is an input vector of 11 elements;
IW is a matrix of weight coefficients of synapses (connections of the first layer with inputs);
b1 is the displacement vector of the first layer of neurons;
b2 is the displacement vector of the output layer;
f(x) is the logarithmic sigmoidal activation function for all neurons in the network;
Y is the output vector of neural network signals (consists of 6 elements);
S1,S2 are the arguments for the activation functions of layer 1 and layer 2, respectively.
Conclusions on the third chapter:
1) The recognition problem using the multilayer perceptron model may have a correct solution. The difficulties of implementing and training a neural network may arise during the formation of initial data. To start the recognition procedure, it is necessary to pre-process the image and decompose each element of the face into separate components.
2) The multilayer perceptron model turned out to be adequate for the selected testing system. In addition, the multilayer perceptron can use different learning algorithms, thanks to its flexible structure and high learning ability.
3) The emotion recognition problem was solved according to a given algorithm.
4) The MATLAB and MathCAD software packages have a wide range of tools that help to implement a mathematical model of a problem, as well as help to train and configure a neural network. These software products quickly calculate cumbersome expressions, which allows you to speed up the process of getting a response.

Conclusion

The implementation of an intelligent system capable of recognizing human emotions is a solution to a technically and mathematically complex problem that requires in–depth study of the subject area, own analysis, skills in processing and selecting experimental data, in-depth knowledge in the field of discrete mathematics, geometry, programming, as well as psychology and many other fields of science.

The use of neural networks in the development of recognition systems allows you to structure poorly formalized data, speed up the data processing process, and assess the adequacy of the environment in which the information model of the neural network is placed.

The metric system allows you to create prototypes of emotions, thanks to which the developer has the opportunity to ”interact” with the neural network, i.e. learning based on the identified prototypes of emotions.

The main purpose of the master’s thesis was to select an information model of a neural network, as well as to describe a recognition algorithm for solving a practical problem.

An information model based on a multilayer perceptron was chosen. This model is optimal in terms of the internal structure and method of managing information flows between neurons. In addition, such a model is able to minimize the number of input elements. The model based on a multilayer perceptron is a universal model and is suitable for solving problems of various levels of complexity, including solving the problem of emotion recognition.

The main conclusions and results of the study:

1) Based on the studied material on artificial intelligent systems, the most effective systems and accurate recognition methods were identified, and successful implementation examples were noted. The system used in this work was self-learning.

2) A review of the main types of information models of neural networks allowed us to identify two universal models applicable to a wide range of tasks: radial neural networks of direct propagation and unidirectional multilayer networks. The choice of an information model for solving a practical problem was made in favor of unidirectional multilayer networks (the multilayer perceptron model). This model is suitable for the internal structure and mathematical description and provides different approaches to solving the problem.

3) The existing methods of emotion recognition in the master’s thesis were considered theoretically. Among the studied recognition methods, a method using artificial neural networks was highlighted.

4) In the practical part of the work, the mathematical apparatus of implementing the recognition algorithm was considered: the choice of face invariants, adequate metrics, pattern recognition by a neural network. A technique for recognizing facial motor activity was also presented and the problem of recognizing emotions in MATLAB and MathCAD data analysis and processing software products was solved.

Based on the results obtained, it can be concluded that the selected information model of an artificial neural network corresponds to the model of a real system and takes into account all its characteristics.

All the tasks were completed in full, therefore, the main goal was achieved.

The experimental part of the work focuses on comparing different face and emotion recognition algorithms implemented in Python using OpenCV library and MediaPipe framework. The aim of the experiment is to evaluate their performance and accuracy in order to select the most suitable algorithm, which will then be integrated into a React JS-based web application. The experiment analyzes the performance of each of the algorithms, evaluating parameters such as image processing speed, accuracy of face and emotion recognition, and robustness to different lighting conditions and face positions. This approach allows not only to identify the most effective technical solutions, but also to take into account the requirements for the user interface and interaction with the end user in the web application.

The selection of the best algorithm to implement in a web application depends on many factors including ease of integration, scalability, and resource intensity. After collecting and analyzing the data, the experimental results provide an informed decision on which algorithm is best suited for use in a dynamic and interactive web application environment where user experience and responsiveness play a key role.

Bibliography

1. Krisilov V.A. Representation of source data in neural network programming problems / Odessa: ONPU. 2003.

2.Levan D.N., Feoktistov N.A. Features of using a multilayer perceptron/ Science Studies. issue 2. 2014.

3. Database and generator of educational resources [Electronic resource]// Bauman Moscow State Technical University, Department of CAD. 2003-2015. URL: http://bigor.bmstu.ru/?cnt/?doc=NN/base.cou (accessed 11.03.2023).

4. Shitikov V.K., Rosenberg G.S., Zinchenko T.D. Methods of system identification / Tolyatti: IEVB RAS. 2003. 463 p.

5. Golovko V.A. Neural networks: learning, organization and application / M.: IPRZHR. 2008

6. Ostroukh A.V. Intelligent systems Textbook/ Scientific and Innovation Center. 2015

7. Profiling [Electronic resource]// Classifications of emotions/ URL:https://ru.wikipedia.org/ / (accessed 01/15/2024).

8. Facial movement coding system [Electronic resource]// basic facial expressions/ URL: http://dic.academic.ru/dic.nsf/ruwiki/1628394 (accessed 02/16/24)

9. Knyazev B.A., Gapanyuk Yu.E. Recognition of abnormal human behavior/ Engineering magazine.2013 512 p

10. IT Galaxy [Electronic resource]// URL: https://ru.intel.com (accessed 02/15/2016).

11. Kozin N. E., Fursov V. A. Step-by-step training of radial neural networks / RFBR. amer. citizen's fund. Published in 2003.

12. Kochetkova A.S. Application of neural networks for security monitoring // Series 9: Research by young scientists. 2007.

13. Zaentsev I.V. Neural networks: basic models/Textbook of physics. F-that. Voronezh. 1999.

14. Krug P.G. Neural networks and neurocomputers: A textbook on the course "Microprocessors" / Publishing House of MEI, 2009.176 p.

15. Kruglov V.V. Fuzzy logic and artificial neural networks / PHYS. 2021.

16. Gradoselskaya G.V. Network measurements in sociology: Textbook / M.: Publishing house "New textbook", 2004. 248 p.

17. Loskutov A.Y. Time series analysis. Course of lectures / Moscow State University. 2023.

18. Lutsenko A.V. Green's function and its application: a methodological guide for the course "Differential Equations" / V.N. Karazin National University of Economics. 2023. 28c.

19. Medvedev V.S., Potemkin V.G. Neural networks in MATLAB / M.:DIALOG-MEPhI. 2002. 496 p.

20. Mestersky L. M. Mathematical methods of pattern recognition / Moscow State University, lecture course. 2004.

21. Minsky M., Papert S. Perceptrons / M.:Mir. 1971.

22. Nazarov A.V., Loskutov A.I. Neural network algorithms for forecasting and optimization of systems / St. Petersburg: Nauka i tekhnika. 2005. 384 p.

23. Nguyen D. A comprehensive study of the classification problem using fuzzy models and distributed computing: dissertation of the Candidate of Technical Sciences. Irkutsk. 2014.

24. Osovsky S. Neural networks for information processing / M.: Williams. 2006.

25. Osovsky S. Neural networks for information processing / translated from Polish by I.D. Rudinsky. 2002. 344 p.

26. Osovsky S. Pattern recognition Overview lecture./M.: Williams. 2007.

27. Ivanov A.I. Neural network algorithms of biometric identification / Publishing house: Radio Engineering. 2006. 144 p.

28. Russell S., Norvig P. Artificial intelligence: a modern approach / Publisher: Williams. 2006. 1424 p.

29. Rosenblatt F. Principles of neurodynamics. Perceptrons and the theory of brain mechanisms / M.: Mir. 1992.

30. Rudakov A.S. Approaches to solving the problem of forecasting time series using neural networks // Data analysis and intelligent systems. No.4. 2018.

31. Rutkovskaya D., Pilinsky, M., Neural networks, genetic algorithms and fuzzy systems / M., 2014. 452 p.

32. Semeykin V. D., Skupchenko A.V. Modeling of artificial neural networks in the MATLAB environment // System. telecommunications. 2008.

33. Soifer V.A. Methods of computer image processing / M.: Fizmatlit. 2001. 784 p.

34. Soldatova O.P., Semenov V. V. Application of neural networks for solving forecasting problems / SSAU named after S.P. Korolev. 2020.

35. Tarkhov D. A. Neural networks. Models and algorithms / Publishing house: Radio Engineering. 2017. 256 p.

36. Tatarenkov D.A. Analysis of methods for detecting faces in an image//Technical sciences. A young scientist. 2015. №4(**??**).

37. Wasserman F. Neurocomputer technology: theory and practice / translated from English by Yu.A.Zuev, V.A. Tochenov. 1992.

38. Khaykin S. Neural networks. Full course / M.: Williams, 2006.

39. Khalafyan A.A. Statistical data analysis. 3rd ed. of studies/ Binom – Press. 2017. 512 p.

40. Tsaregorodtsev V. G. Computing technologies// Bulletin of KazNU. ch 3. 2018. pp. 308-315.

41. Shamis A.L. Behavior, perception, thinking: problems of creating artificial intelligence / Publishing house: Nauka. 2015. 224 p.

42. Shakhnov V.A., Vlasov A.I., Kuznetsov A.S. Neurocomputers: architecture and circuit engineering / M.: Publishing House of Mechanical Engineering. 2020.

64 p.
43. Efremova E.A., Dunaev E.V. Application of neural networks for forecasting financial time series.//Reports of TUSUR. 2004.
44. Kashirina I. L. Neural network technologies: An educational and methodological guide for universities / VSU Publishing House. 2008. 72 p.
45. Analytical technologies for forecasting and data analysis// Neuroproject. 1999-2005.[Electronic resource]/URL: http://www.neuroproject.ru/neuro.php (accessed 11.02.2016).
46. Gundyrev K.V. Artificial neural networks in the tasks of diagnosing rail circuits// Scientific research.the laboratory. "Computer automation systems". 2005.
47. Introduction to the theory of neural networks [Electronic resource]//Basic concepts of neural networks/ URLs: http://www.orc.ru (accessed 12.04.2023)
48. Market research of emotion recognition systems for networks [Electronic resource]/ URL: https://habrahabr.ru/post/133686 / (accessed 2.02.2024).
49. Areas of application of artificial neural networks [Electronic resource]//The main directions of use/ URL: http://www.neuropro.ru / (accessed 02/19/2024)
50. The concept of information [Electronic resource]./ URL: http://cde.osu.ru / (accessed 12.02.2024)
51. Artificial Intelligence Portal [Electronic resource]// Artificial neural networks /URL: http://www.aiportal.ru (accessed 01.02.2024)
52. Application of neural networks in robotics [Electronic resource]/URL: http://geektimes.ru / (accessed 03.12.2023)
53. Application of neural networks for classification tasks [Electronic resource]/URL: http://www.basegroup.ru / (accessed 02/5/2024)
54. Callan R. Basic concepts of neural networks/ Ed.: Williams. 2002. 287 p
55. Development and use of intelligent systems [Electronic resource]/ URL: http://www.studfiles.ru/preview/5809976 / (accessed 14.01.2024)
56. Kalchenko D.A. Neural networks on the threshold of the future. / Computpress, 2005.
57. Emotional tension determination methodology [Electronic resource]/URL: http://pandia.ru/text/80/079/8545.php (accessed 02/19/2024)
58. MATLAB [Electronic resource]//introduction to the program/ URL: http://matlab.ru/pro
(accessed 02/29/2024)
59. MATLAB. [Electronic resource]// Laboratory workshop/URL: http://elib.rshu.ru/files_bo
503135146.pdf (accessed 2.05.2023)
60. MATLAB. [Electronic resource]// Infowork /URL :https://infourok.ru (accessed 01/10/2023)
61. Microsoft Excel [Electronic resource:]// Editing tables /URLs: https://office.live.com/star
RU (accessed 01/17/2023)