CATEGORICAL ASSESSMENT OF DEPRESSION BASED ON LOW LEVEL FEATURES

by

MARIDAKI M. ANNA

BSc. Technological Educational Institute of Central Macedonia, 2013

A THESIS

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

DEPARTMENT OF INFORMATICS

SCHOOL OF ENGINEERING

TECHNOLOGICAL EDUCATIONAL INSTITUTE OF CRETE

Approved by:

Professor Manolis Tsiknakis

Heraklion, Crete

2018

Statement of Originality

I hereby certify that I am the sole author of my thesis entitled "*Categorical Assessment of Depression Based on Low Level Features*". This is the original work performed by me under the guidance and advice of my faculty advisor, Manolis Tsiknakis and the Graduate Studies office of Technological Educational Institute of Crete. The research work contained in this thesis was conducted between October 2016 and March 2018. It is original work except where due reference is made. It has not been submitted for the award of any degree or diploma to any other institution of higher learning.

Anna Maridaki

May 2018

Abstract

Mental illness is a disease which usually causes behavior disturbances. Nowadays, mental disorders are commonplace, affecting a large number of people. There are more than 200 forms of mental illness. Depression, bipolar disorder, dysthymia and anxiety disorder are some of the most common mood disorders that are inherently related to emotions. According to the European Statistical System [1], 7.1% of the European citizens reported having chronic depression. Whereas, the World Health Organization ranks major depressive disorder as the 4th leading cause of disability worldwide. Major depressive disorder (MDD), also known as clinical depression, is a mood disorder involving bad mood, low self-esteem and loss of interest in normal pleasurable activities. Depression affects the way of thinking, feeling and acting in daily duties.

Major depressive disorder has observable behavioral symptoms. Actually, some behavioral cues are closely associated with depression, such as body language, speech, head and face movement. In most cases, facial expressions indicate individual’s depressed feelings. Technology has the potential to assess those cues for depression diagnosis. There are numerous research groups focused on automated depression detection based on audio and visual signals analysis. Automatic recognition of human emotions have a lot of applications in Human Computer Interaction and Affective Computing field. However until now the performance of such applications is still not satisfying.

The aim of this work is to develop a framework for the assessment of major depressive disorder, as a supportive application for the clinical care of patients. In so doing we utilize the dataset of the depression recognition sub-challenge of the Audiovisual Emotion Recognition challenge (AVEC), as it is the only publically available dataset consisted by video recordings that are annotated with a depression index. Algorithmic methods were developed, using MATLAB, for the detection of depression using low level image-based features. In the proposed framework, two different motion representation methods were tested, namely: a) Motion History Image (MHI), and b) Gabor Motion History Image (GMHI). Appearance-based descriptors employed were the employed, namely Local Binary Pattern (LBP), Local Phase Quantization (LPQ), and Histogram of Oriented Gradients (HOG). Additionally, some statistical features were utilized. Subsequently, machine learning algorithms were selected in order to define the specificity of the proposed framework. The selected classification algorithms are k-Nearest Neighbors (kNN), Random

Forest, Support Vector Machine (SVM) and Naïve Bayes. For the MHI approach the best performance 81.93% achieved by appearance-based descriptor HOG, and with combination of HIST-MEAN-STD using SVM classifier for 100 selected features. The execution time for those results are 1.46 seconds and 0.15 seconds respectively. As for GMHI approach the maximum F1 score is 81.93% achieved by the combination of statistical features HIST-MEAN-STD for 10 selected features and the execution time is 0.12 seconds.

Keywords: Depression Assessment, Major Depressive Disorder (MDD), AVEC 2014 Dataset, Low Level Features, Principal Component Analysis (PCA), Image Processing, Facial Image Analysis, Motion History Image (MHI), Gabor Inhibition

Περίληψη

Η ψυχική ασθένεια είναι μια ασθένεια η οποία συνήθως προκαλεί διαταραχές στη συμπεριφορά. Στην εποχή μας, οι ψυχικές διαταραχές είναι συνηθισμένες, επηρεάζοντας ένα μεγάλο αριθμό ανθρώπων. Υπάρχουν περισσότερες από 200 μορφές ψυχικής ασθένειας. Η κατάθλιψη, η διπολική διαταραχή, η δυσθυμία και η διαταραχή άγχους είναι μερικές από τις συνήθεις διαταραχές της διάθεσης που σχετίζονται εγγενώς με τα συναισθήματα. Σύμφωνα με το Ευρωπαϊκό Στατιστικό Σύστημα [1], το 7,1% των ευρωπαίων πολιτών ανέφερε ότι πάσχει από χρόνια κατάθλιψη. Ενώ, ο Παγκόσμιος Οργανισμός Υγείας κατατάσσει τη Μείζονα Καταθλιπτική Διαταραχή ως την 4η κύρια αιτία αναπηρίας παγκοσμίως. Η Μείζονα Καταθλιπτική Διαταραχή (MDD), επίσης γνωστή ως κλινική κατάθλιψη, είναι μια διαταραχή της διάθεσης που περιλαμβάνει κακή διάθεση, χαμηλή αυτοεκτίμηση και απώλεια ενδιαφέροντος για φυσιολογικές ευχάριστες δραστηριότητες. Η κατάθλιψη επηρεάζει τον τρόπο σκέψης, τα συναισθήματα και τη συμπεριφορά σε καθημερινά καθήκοντα.

Η Μείζονα Καταθλιπτική Διαταραχή έχει παρατηρήσιμα συμπεριφορικά συμπτώματα. Στην πραγματικότητα, ορισμένα συμπεριφορικά σημάδια σχετίζονται στενά με την κατάθλιψη, όπως η γλώσσα του σώματος, η ομιλία, η κίνηση του κεφαλιού και του προσώπου. Στις περισσότερες περιπτώσεις, οι εκφράσεις του προσώπου υποδηλώνουν τα καταθλιπτικά συναισθήματα του ατόμου. Υπάρχουν πολυάριθμες ερευνητικές ομάδες που επικεντρώνονται στην αυτοματοποιημένη ανίχνευση της κατάθλιψης βασισμένη στην ανάλυση ακουστικών και οπτικών σημάτων. Η αυτόματη αναγνώριση των ανθρώπινων συναισθημάτων έχει πολλές εφαρμογές στον τομέα αλληλεπίδρασης ανθρώπου υπολογιστή και της συναισθηματικής πληροφορικής. Ωστόσο, μέχρι στιγμής η απόδοση τέτοιων εφαρμογών δεν είναι ακόμα ικανοποιητική.

Ο στόχος αυτής της εργασίας είναι να αναπτυχθεί ένα πλαίσιο για την αξιολόγηση της μείζονος καταθλιπτικής διαταραχής, ως υποστηρικτική εφαρμογή για την κλινική φροντίδα των ασθενών. Για να γίνει αυτό, χρησιμοποιούμε το σύνολο των δεδομένων για την αναγνώριση της κατάθλιψης από τον διαγωνισμό Οπτικοακουστική Αναγνώριση Συναισθημάτων (AVEC), καθώς αποτελεί τη μοναδική βάση δεδομένων που διατίθεται στο κοινό, αποτελούμενη από οπτικοακουστικό υλικό συνοδευόμενο από ετικέτες με τον αντίστοιχο δείκτη κατάθλιψης. Αλγοριθμικές μέθοδοι αναπτύχθηκαν, κάνοντας χρήση του λογισμικού MATLAB, για την ανίχνευση της κατάθλιψης χρησιμοποιώντας χαμηλού επιπέδου χαρακτηριστικά βασισμένα σε εικόνα. Στο προτεινόμενο πλαίσιο δοκιμάστηκαν δύο διαφορετικές μέθοδοι αναπαράστασης της κίνησης, συγκεκριμένα: α) Motion History Image (MHI) και β) Gabor Motion History Image (GMHI). Οι περιγραφείς με βάση την εμφάνιση που χρησιμοποιήθηκαν, είναι το τοπικό δυαδικό πρότυπο (LBP), η τοπική ποσοτικοποίηση φάσης (LPQ) και το ιστόγραμμα των προσανατολισμένων διαβαθμίσεων (HOG). Επιπρόσθετα, χρησιμοποιήθηκαν ορισμένα στατιστικά στοιχεία. Ακολούθως, επελέγησαν αλγόριθμοι μηχανικής μάθησης για να εξετασθεί η «ειδικότητα» (“specificity”) του προτεινόμενου πλαισίου. Οι επιλεγμένοι αλγόριθμοι ταξινόμησης είναι k-Εγγύτατου Γείτονα (k-Nearest Neighbors), Δένδρα Αποφάσεων (Random Forest), Μηχανή Διανυσμάτων Υποστήριξης (Support Vector Machine) και Naïve Bayes. Για την προσέγγιση MHI η καλύτερη απόδοση ήταν 81,93% που επιτεύχθηκε με τα χαρακτηριστικά HOG και με τον συνδυασμό από τα χαρακτηριστικά HIST-MEAN-STD χρησιμοποιώντας SVM ταξινομητή για 100 επιλεγμένα χαρακτηριστικά. Ο χρόνος εκτέλεσης για αυτά τα αποτελέσματα είναι 1,46 δευτερόλεπτα και 0,15 δευτερόλεπτα αντίστοιχα. Όσον αφορά την προσέγγιση GMHI, το μέγιστο F1 είναι 81,93% και επιτυγχάνεται με τον συνδυασμό των στατιστικών χαρακτηριστικών HIST-MEAN-STD για 10 επιλεγμένα χαρακτηριστικά και ο χρόνος εκτέλεσης είναι 0,12 δευτερόλεπτα.

Λέξεις-κλειδιά: Αξιολόγηση Κατάθλιψης, Μείζονα Καταθλιπτική Διαταραχή (MDD), Σύνολο Δεδομένων AVEC 2014, Χαρακτηριστικά Χαμηλού Επιπέδου, Ανάλυση Κύριων Συνιστωσών (PCA), Επεξεργασία Εικόνας, Ανάλυση Εικόνας Προσώπου, Motion History Image (MHI), Gabor Inhibition

Table of Contents

List of Figures

List of Tables

Acknowledgements

First of all, I would like to express my deep gratitude to Professor Manolis Tsiknakis, supervisor of the present master thesis. Thank you for giving me the opportunity to become a member of your team and for the active supervision, motivation and encouragement.

Besides my supervisor, I would like to express my sincere acknowledgement in the continuous support and guidance of my thesis advisor Anastasia Pampouchidou. It was pleasure to work with you and learn from you. A special appreciation to Dr. Matthew Pediaditis for his constructive suggestions that improved the results obtained in this thesis.

Last but not least, I would like to thank the members of Biomedical Informatics & eHealth Laboratory for their support and pleasant working environment, which made this journey enjoyable.

Anna Maridaki

1. Introduction

Humans are characterized by changes in their affective state. Sometimes these changes may last longer or may impact an individual’s proper functioning. Affective disorders are a set of psychiatric disorders, also called mood disorders. The three main types of affective disorders are depression, bipolar disorder, and anxiety disorder. Each includes subtypes and variations in severity. Depressive Disorder is classified as a mood disorder causing distressing symptoms, affecting feelings and the handling of daily activities. World Health Organization (WHO) enlist depression as the fourth most significant cause of suffering worldwide and estimates that it will be the leading cause by 2020 [2]. An updated statistical estimation [3] reports that 350 million people throughout the world are affected by depression, while European Union [4] estimates that mental health problems affect one in four citizens at some point during their lives. At its worst, depression can lead to suicide. Actually, the risk is 30 times higher for patients affected by depression, than the general population [5]. Close to 800.000 people die due to suicide every year, while suicide is the second leading cause of death in 15-29-year-olds. [6].

Depressive disorder is a common and serious medical illness that negatively affects many aspects of individual’s daily duties. In particular, feelings, way of thinking and acting are influenced. Major depression can lead to a variety of emotional and physical problems and can decrease person’s ability to function in social life. The Diagnostic and Statistical Manual (DSM) of mental disorders [7] describes a range of audiovisual indicators of depression. These include sadness, tears, inability to sit still, pacing, hand-wringing and other signs of psychomotor agitation, slower speech and body movements, reduced interpersonal responsiveness, and decreased vocal intensity and inflection in psychomotor retardation [8].

Despite the scientific studies that have already been conducted, remarkably little innovation has occurred in the clinical care of depressive disorder. The most widespread diagnostic method is based on diagnostic and statistical manual of mental disorder criteria and on questionnaires such as BDI [9] and Hamilton depression rating scale [10]. According to WHO [2] depression detection failure is due to the lack of resources and trained care providers. Even though major depression is one of the most common mental disorders, it is often difficult to be diagnosed, as it manifests in different ways and also because the assessment methods rely on subjective clinical judgment. Currently, there is no objective method to diagnose depression, nor a laboratory-based test.

Affective Computing and Social Signal Processing are two promising fields for a better and more accurate depression assessment. To this direction automated depression detection systems have been proposed in order to support clinicians' decision and avoid mistaken diagnosis. Such systems will also be helpful to overcome the problem of subjective bias associated with self-reports.

Currently, video-based systems for automated depression assessment have only been evaluated in research settings, and have not been applied in the general population in order to evaluate their feasibility [11]. The "Audio/Visual Emotion Challenge" AVEC 13 [12], AVEC 14 [13] and AVEC 16 [14] provide a dataset that can support the development of an automated depression detection system. The provided dataset is the only publically available dataset that includes video recordings annotated with a depression index. This challenge gave the opportunity to the researchers for the development of automated systems, which can objectively compare their performance and set up state of the art approaches.

In the present work we developed an automated framework for categorical depression assessment based on low level, image-based features. The AVEC dataset is used for the development and testing of our computational framework. The proposed work is based on the hypothesis, supported by available scientific evidence, that depressed individuals exhibit restricted facial motion than non-depressed. Face is the region of interest where two motion representation methods are implemented (Motion History Image and Gabor Motion History Image), where moving parts of a video sequence can be engraved in a single image. For feature extraction low level appearance-based descriptors were developed which are Local Binary Pattern, Local Phase Quantization and Histogram of Oriented Gradients. Additionally, same statistical features were utilized. Mean (average value), standard deviation (mean deviation of the signal compared to the average) and combined histogram. For the system evaluation Naïve Bayes, k-Nearest Neighbors, Random Forest and Support Vector Machine classification algorithms were used. The maximum performance is achieved with SVM classifier for both MHI and GMHI motion representation methods. For the MHI approach the best performance 81.93% achieved by appearance-based descriptor HOG and with combination of HIST-MEAN-STD using SVM classifier for 100 selected features. The execution time for those results are 1.46 seconds and 0.15 seconds respectively. As for GMHI approach the maximum F1 score is 81.93% achieved by the combination of statistical features HIST-MEAN-STD for 10 selected features and the execution time is 0.12 seconds.

1.1 Research Questions

In the present work our aim is, as already stated, to develop an automated computational framework for categorical depression assessment based on low level, image-based features . Our attention is focused on the optimal configuration of a set of parameters to maximize system performance with the smallest execution time. The following questions are posed and investigated in this work:

- Which feature vector will maximize the overall system performance?
- Which PCA value optimizes the overall system performance?
- Which of the implemented machine learning classifier will provide the maximum F1-score with the minimum execution time?

The main contribution of the proposed work lies in the introduction a novel variant of motion history image that is Gabor Motion History Image.

1.2 Thesis Outline

This thesis comprises six chapters, including this introduction. Chapter 2 provides the clinical perspective of depression diagnosis and symptoms according to DSM-5. Literature review related to the automated depression diagnosis is included in section 3. Also, the relevant implemented algorithms are presented. Chapter 4, describes the general methodology that was followed in this research to build a depression recognition system. The experimental results, the different configuration parameters and the execution time are analyzed in section 5. The final chapter involves a summary of the contributions of this thesis and the future directions for ongoing research are discussed.

1.3 List of Publications

During this Master three publications have been produced.

i. A. Pampouchidou, O. Simantiraki, C. - M. Vazakopoulou, C. Chatzaki, M. Pediaditis, A. Maridaki, K. Marias, P. Simos, F. Yang, F. Meriaudeau, M. Tsiknakis,"Facial Geometry and Speech Analysis for Depression Detection", 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 11-15 July 2017, Jeju Island, Korea, (1433:1436) [15].

ii. A. Pampouchidou, M. Pediaditis, A. Maridaki, M. Awais, C.-M.Vazakopoulou, S.Sfakianakis, M,Tsiknakis,P. Simos, K. Marias, F. Yang, F. Meriaudeau, "Quantitative comparison of motion history image variants for video-based depression assessment", EURASIP Journal on Image and Video Processing, 2017. (1), 64 [16].

iii. A. Maridaki, A. Pampouchidou, K. Marias, M. Tsiknakis, "Machine Learning Techniques for Automatic Depression Assessment", 41st IEEE International Conference on Telecommunications and Signal Processing, July 4-6 2018, Athens, Greece (Accepted to be presented)

2. Clinical Depression Analysis

Depression is a mood disorder that causes intense symptoms affecting feelings, thoughts, and the ability to handle daily activities. As reported by the World Health Organization depression is the most frequent mental disorder with more than 300 million affected individual worldwide [6]. Major depressive disorder, disruptive mood dysregulation disorder, premenstrual dysphoric disorder, persistent depressive disorder (dysthymia), substance/medication-induced depressive disorder, depressive disorder due to another medical condition, and other specified depressive disorder and unspecified depressive disorder are included in the term of depressive disorder. All the above mentioned disorders have some mutual characteristics such as depressed mood joined by physical and cognitive changes that fundamentally influence the person's pleasure in most activities and hobbies. The things that varies among the above mentioned disorders are the presumed etiology, the duration and the timing [17].

The most common form of depression is Major Depressive Disorder (MDD). It is characterized by depressive symptoms most of the day, nearly every day or for at least two weeks duration. On the other hand, persistent depressive disorder (dysthymia) is a more permanent and severe form of depression [17]. Disruptive mood dysregulation disorder (DMDD) is a mental disorder in children and adolescents characterized by extreme irritability, anger, intense temper outbursts as well as severe daily impairment in school or with friends, that requires clinical attention [18]. Another form of depression is premenstrual dysphoric depression disorder (PMDD). It is about a severe, chronic and disabling form of premenstrual syndrome which symptoms affect women one to two weeks premenstrually. Symptoms of PMDD include irritability depressed mood, anxiety and marked somatic and behavioral changes [19]. Certain medications, drugs of abuse or withdrawal can lead to substance/ medication –induced depressive disorder. Other specified depressive disorder and unspecified depressive disorder are categories used when depressive symptoms cause problems with school, work, relationships with others, or daily activities. Certain medical condition can cause depressive disorder due to another medical condition, which is a subtype of depression.

2.1 Tools for Depression Assessment

Depression diagnosis is not based on dedicated laboratory procedure; on the contrary it is part of an integrated mental health estimation. Evaluating past and current symptoms, medical and family history in conjunction with professional observation will end up to a complete diagnosis. One of the most widely used rating scales for depression has been proposed in 1960 by Hamilton [10] [20]. The Hamilton Depression Rating Scale, also called Hamilton Rating Scale for Depression (HRSD), is a psychiatric measuring instrument used to indicate the severity of depression. The questionnaire includes 21 questions, but only 17 are used in scoring. Depression severity is measured by probing mood, anxiety, feelings of guilt, agitation or retardation, insomnia, suicide ideation, weight loss, and somatic symptoms [21]. Between 3 and 5 are the possible responses of the question sheet. Clinicians conclude to the diagnosis after following a structured interview with the patient and after observing the patient's symptoms [20].

There are also scales fully completed by patients. They are types of self-report rating inventories that a person accomplish without the professional assistance such as Beck Depression Inventory (BDI) [9]. It is one of the most widely used questionnaire measuring depression severity, consisted of 21 multiple-choice questions which represent categories of symptoms and attributes. Each category describes a specific behavioral manifestation of depression. The categories are mood, irritability, pessimism, social withdrawal, sense of failure, indecisiveness, lack of satisfaction, body image, work inhibition, guilty feeling, sleep disturbance, sense of punishment, fatigability, loss of appetite, self-hate, weight loss, self-accusations, somatic preoccupation, self-punitive wishes, loss of libido and crying spells [9].

The Diagnostic and Statistical Manual of Mental Disorders (DSM) is published by the American Psychiatric Association (APA) and is widely used by health care professionals as a formal consulting guide for mental disorder diagnosis [22]. It involves symptoms descriptions and criteria diagnosis for the classification of mental disorders. The first edition was released in 1844 [17], since then several editions followed. In 2013 the last edition was published, which has replaced the DSM fourth edition, after a decade of research and contains the most updated criteria for diagnosing mental disorders [22].

2.2 Major Depressive Disorder

Sadness is a natural part of human experience. Feeling down, empty and useless, having loss of pleasure and interest in every daily activity are familiar symptoms to all. However, if such feelings firmly continue and affect individual’s life substantially then, it might be depression. Major Depressive Disorder (MDD) is more than just feeling sad or fatigue. Actually, it is a serious mental health condition that in most cases requires clinical monitoring and medical care. Major depressive disorder, frequently referred to simply as depression or clinical depression, is a common and serious medical illness that negatively affects many areas of individual’s life. In particular, feelings, way of thinking and acting are influenced. It impacts mood and behavior as well as various physical functions, such as appetite and sleep.

2.2.1 Symptoms of Major Depressive Disorder

Depressed mood or loss of pleasure most of the day accompanied with loss of interest in activities once enjoyed for at least two weeks duration is the essential diagnostic feature of major depressive disorder. Affected individuals in childhood or adolescence is more likely to have irritable mood rather than feelings of depression [17]. In order to characterize an individual as affected by major depressive disorder, at least four additional symptoms should be present. Those can be diminished ability of thinking or indecisiveness, increase or decrease in appetite, excessive or moderate sleeping hours, psychomotor agitation or retardation, worthlessness feeling and recurrent suicidal thoughts. Major depressive episodes last at least two consecutive weeks and their symptoms affect patients during the day, nearly every day.

Trying to describe the main and the most significant symptom, people express their feelings as sad, disappointed and hopeless. Increased irritable mood, persistent anger, blaming others and unreasonable frustration are common diagnostic features of major depressive disorder. In most cases, facial expressions indicate individual’s depressed feelings. In particular he/she looks as if he/she is about to cry [17]. The presence of a depressed mood can be inferred from the person's facial expressions, which are remarkably reduced. Furthermore people suffering from depression complain about body pain and aches.

Major depression can lead to a variety of emotional and physical problems, and can decrease a person’s ability to function at work and at home. People suffering from MDD complain they feel so tired that they do not have energy for their daily activities even when they sleep or rest long. The clinical condition of same patients force them unable to experience pleasure from usually enjoyable activities and hobbies. It is a common pattern for, affected individuals to skip activities they normally enjoy and isolate themselves from the world. Social withdrawal is a common symptom for people suffering from major depressive disorder.

Depression is a mental illness, but it can affect your body as well as your mind. Sleeping problems, for example, can be a symptom of depression. The difficulty in falling asleep, as well as any restless sleep with many interruptions are usual symptoms. On the other hand, some people can get to the other end and sleep excessive hours a day. Insomnia is a common depressive sign characterized by the difficulty of relaxation during sleep hours [17]. Hypersomnia, on the contrary, is characterized by prolonged sleep, excessive sleepiness and too deep sleep.

Depression appears to be related with nutritional behavior. Affected adults usually describe their eating habits as they forced to eat or as they do not have desire to eat. Poor appetite is not the only appetite change. Change of appetite can also refer to a significant increase in appetite, binge eating, or cravings for specific foods. Appetite changes can result in a significant weight loss or weight gain. In terms of children, significant weight loss is considered the failure of weight gaining according to the standard growth charts [17].

Psychomotor symptoms are also related to major depressive disorder. Agitation such as pacing, hand-writing, cloth or skin pulling, and the inability to remain still are diagnostic criteria of the disorder [17]. Slowing-down of thought and reduction of physical movements in an individual is also observed. For example, slowed verbal expression, meditation, and changes of physical position, increased pauses before answering and low speech intensity are relevant psychomotor symptoms of depression. According to DSM-5 [17], “the psychomotor agitation or retardation must be severe enough to be observable by others and not represent merely subjective feelings”.

Along with making you feel very sad, depression can also make you feel drained of energy. Being tired all the time is a popular complaint. Physical or mental exhaustion and fatigue are common problems. Tiredness can negatively impact performance at work, family life, and social relationships. Even daily routine activities seem to require substantial effort and may cause feelings of being tired. For instance, it is common washing and dressing takes much longer than usual and make patients feel wash out with reduced energy [17].

Some people are just more prone to negative thinking, while others set impossibly high standards for themselves. Sense of worthlessness is a common feeling of major depressive episode. Many individuals report feeling difficult to live up to other people's expectations of them, or to their own expectations. They report an exaggerated sense of culpable of every failure or trivial events even if they are not responsible for them. For example they strongly believe that worldwide abject poverty is their responsibility [17].

When someone is under the influence of major depression disorder, his or her mind is overstimulated or distracted. They may also find that they cannot think clearly and focus or maintain their attention on a task, which may affect their decision-making and work or school performance. The inability to remember become apart for the individual's day to day experience. In some cases experiences of forgetfulness in elderly individuals may be correlated with early signs of dementia ("pseudodemendia") [17]. The ability to remember information and events someone would normally be able to recall may be fully abated, if major depressive disorder prosperously treated. Yet, there are cases where a depressive episode may be the initial presentation of an irreversible dementia.

Experiencing difficult life phases, such as depressive episodes someone can be led in negative thoughts and intense feelings of despair. Thoughts such as "I wish I died to rest" reflect the experience of the suffering and the deadlock they experience because of their external adversities and internal weaknesses. In most severe circumstances, affected individuals put their pending in order. For example, they try to solve misunderstandings, update wills or even plan for carrying out the whole suicidal process. The patient is biologically or psychologically vulnerable, several factors trigger or contribute to suicidal behavior at a specific time. The one who really seeks to commit suicide believes that he has considered all the practical possibilities of getting out of pain without result. Believes that suicide is the only way to stop suffering. Suicide appears to be a unique possibility of redemption from troubles, as an act of violent exodus from unstoppable mental pain [17].

2.2.2 Difference between Bereavement and Depression

Normal sadness such as bereavement and depression share some common symptoms thus, clinicians run the risk of confusing depression diagnosis with sadness. Bereavement affects people in different ways. Some of the symptoms that arise from losing someone are similar to depression, such as withdrawal from social settings and intense feelings of sorrow. Prolonged feelings of sadness can lead to depression or make underlying depression worse. However, there are important differences between depression and grief. In grief the predominant feeling is emptiness and loss. On the other hand, depressed mood consists affected individuals unable to anticipate pleasure or happiness. Major depressive episodes are more persistent and intense in contrast to grief that occurs in waves. Such waves are called pangs of grief and are likely to decrease their intensity over days to weeks. Humor and positive emotions are not characteristics of depression still, grief may be accompanied with such feelings. As far as the thought content, in grief is associated with the passed away person but in depression is more pessimistic and self-critical. Major depressive episodes affects individual's self-confidence. They feel worthless and have a pessimistic or negative view of their self, whereas in grief self-esteem is preserved. Thoughts about death and dying is common in both situations. Bereaved individual's ideation is focused on the possibility of joining the deceased while people suffering from depression thing they are unable to cope with pain of depression and have suicidal ideations [17].

2.2.3 Etiology – Contributing Factors

There is a set of causes that trigger the onset of depression. It is a combination of environmental, physiological and social factors. Depression is likely to be developed at any age. However, adolescents are more vulnerable. As stated by the American psychological association in the United States, the 20s is more prone than other ages. Actually, the prevalence in 18- to 29-year-old individuals is threefold higher than the prevalence in 60-year–old individuals [17]. There are also differences between genders in prevalence rates for depressive disorders. As a matter of fact, female population manifest 1.5-to 3-times higher rates than male population in early puberty [17].

Depression is an inherited condition thus some people are at an increased genetic risk. Having a first-degree family member suffering from major depressive disorder is two-to four fold higher to expose depressive episode at some point in their life than the general population [17]. There are environmental factors that influence individual's mood. Individuals are affected by outside events differently. Depression can be triggered by adverse life situations or other factors such as health issues, especially when the person has a chronic health problem, adverse childhood experiences and bereavement. Life circumstances and other personal factors are likely to have an important influence. Close relationship between depression and other chronic diseases also exist. Medical conditions increase the danger of developing major depression episode. In this case, these episodes often follow a more refractory course than the depressive episodes in medically healthy individuals. Morbid obesity, Diabetes, and cardiovascular disease are responsible illnesses for developing depression [17].

2.3 Nonverbal Cues of Depression

Communication is involved in all humans' life. The ability to communicate at an advanced level distinguishes human beings from animals. Human communication and interaction is based on verbal and nonverbal signs. Interpersonal communication may be verbal using a spoken language and non-verbal, expressed by facial expressions or gestures, body movements or postures, head pose, movement as well as eye gaze and blinking. The interpretation of nonverbal cues is as significant as verbal since it contains information regarding emotions [23]. "Facial Expression is one of the most powerful, natural and immediate means for human beings to communicate their emotions and intentions" [24]. According to Ekman [25] face observation may draw conclusions about age, gender as well as what the person thinks or feels. Not surprisingly, face region has been of keen interest to depression detection.

Identifying the effects of depression on facial expression is a vigorous and early work research area. The importance of nonverbal cues in therapeutic interaction was declared by McGowan and Schmidt. They claimed that "it may even be that one of the major variables which will help us to distinguish the work of the experienced counsellor from that of the novice, is his ability to pick up and respond to minimal nonverbal cues in an accurate manner" [26]. Waxer et al. [27] report that eyes, mouth and angle of head are nonverbal cue areas for depression. In the early work [28] found that facial expression of depressed people are unlike to those of non-depressed individuals. Facial expressions hold effective cues in diagnosing depression.

Head movement reflect cues about mood and emotions. In the opinion of Waxer [29] depressed people is more likely to position their head downward than non-depressed and also avoid doing head movements or nodding. Depression in general, causes slowing body and facial movement. Even eye glances and brow raising can be a data source for the mental health evaluation such as the depression level [30]. Facial expressions of depressed individuals have been extensively examined. According to Ekman [31] the principal universal emotions are joy, surprise, anger, sadness, disgust, and fear. Yet, detailed investigations [32] stated that depressed look, sad, and disgust are more often than non-depressed. The most dominant emotion is intense sadness [33] with long lasting sadness periods. Also, affected individual have more unfelt smile and less felt smile [32]. Other studies focus on eyes' movement, such as Alghowinem et al. [34] who concluded that the average duration of blinks is significantly longer in depressed compared to not-depressed individuals. Smile duration and intensity was also found to be a diagnostic feature of depression too [35]. When it comes to depression assessment, the whole face and individual's facial expressions and movements are powerful indicators.

3. Literature Review

Emotional experience does not solely occur in mind. It can be inferred using many different modalities such as change body postures, biophysiological measurements as well as cues for facial expressions. Estimation of the affective state is the main scope of a novel subfield of computer science called Affective Computing. Affective computing is an active and interdisciplinary research area, with several fields having contributed to it during the past decades, including computer vision, machine learning, and psychology. This growing field endeavors to give electronic devices emotional intelligence so that they can respond to human feelings [36].

As depression affects mental health it has strong correlation with feelings. Motion face analysis plays significant role whereas, detecting face and facial features has become an important task. In this thesis based on image processing techniques, facial feature detection algorithms are developed. According to Phimoltares et al. [37] there are five categories of face and facial feature detection. Those are geometry-based methods, color-based approaches, appearance-based methods, motion-based methods and edge-based methods. Another taxonomy proposed is geometry-based and appearance-based [38], [39]. Geometric features utilize distances and shapes and other geometric properties whereas appearance use gray value (intensity) information of the image [37]. An additional feature taxonomy is low and high level [40]. High level features can be interpreted by humans in contrast with low level that are detected by image processing algorithms and do not correspond to common sense knowledge. In the present work appearance-based low level features were utilized for the full face region.

Within the past decades different ways of emotion recognition are in the focus of numerous research groups all over the world. Since the investigators may use different datasets, contexts and develop a wide variety of methods, these approaches are not always objectively comparable. Great importance related work has been done on automatic depression prediction and state-of-the-art approaches has been achieved. Some research groups focus their experiments using speech and prosody [41], [42] other head pose, gesture and facial expressions [43], [44], [45] as well as multimodal combination of these cues has been done [46], [15], [47], [48].

3.1 Automated Depression Analysis

Depression is characterized by strong observable indicators related with general psychomotor functioning, such as facial expressions, body and head movements as well as other nonverbal cues. Primarily depression diagnosis and assessment of symptom severity rely on subjective measures of behavior. However, computational approaches gave the opportunity for automated depression analysis that human hardly manage to quantify.

Computer scientists in early 70s concerned about using face as a potential biometric indicator [49], and decades later it was utilized in computer vision for automated depression analysis [50]. In line with [51] and [11], several researchers follow the same structure of analysis (Figure 1). Having a visual signal, the first step is the preprocessing where the relevant body parts (e.g., face, head or torso) are detected within the image sequence. Then, the feature extraction step follows, applying feature extraction algorithms at the detected regions according to the research question. Subsequently, feature selection methods are implemented in order to reduce feature vector high dimensionality. The final step is classification and regression, where machine learning algorithms are developed to interpret the features.

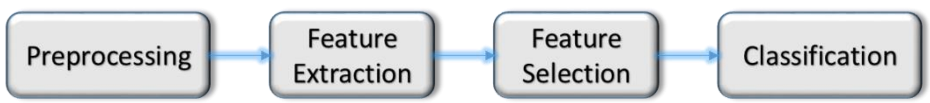


Figure 1 General execution pipeline of automated depression analysis

3.2 Motion Representation Methods in Depression Assessment

The present work focuses on facial expression when it comes to depression assessment, as people express their emotions through the visual signal. Several attempts have been made to model the mapping between facial expressions and emotional states [52]. The basic emotions according to Ekman (Figure 2) are joy, surprise, anger, sadness, disgust, and fear [31]. Some expressions are subtle and it is impossible to deduce the emotional state of a person from a still face image. Knowing that affected individuals exhibit less motion in their facial expressions than healthy individuals, implemented methods derived from Motion History Image (MHI). This technique can represent a motion sequence in compact manner, in a scalar-valued image [53]. The algorithm depicts the most recent motion with white pixels. Gray color corresponds to the former motion and black pixel interpreted as immovability. Actually, it is a static representation of motion showing where motion occurred. The brighter gray the most recent motion.

Motion History Image (MHI) algorithm is commonly used in the field of human action recognition [54], as it is a simple algorithm, easily implemented with good performance [55]. More recent evidence shows that Motion History Image was implemented to recognize human facial

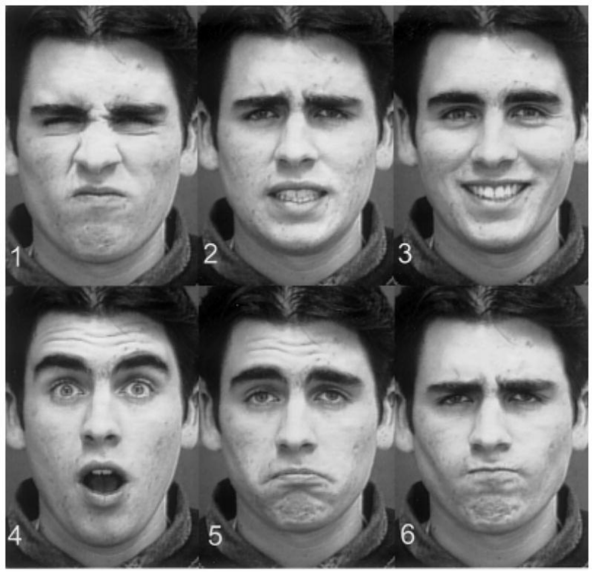


Figure 2 Basic facial expression 1 disgust, 2 fear, 3 joy. 4 surprise, 5 sadness, 6 anger [109]

expressions [56]. Related literature includes Motion History Histogram (MHH) algorithm which is an extension of MHI. Meng el al [57] draw attention to a multimodal automatic depression recognition method, validated on the AVEC2013 dataset. MHH was suggested for capturing the motion of each pixel within the face, describing temporal information during facial expressions. Edge Orientation Histogram (EOH) and Local Binary Patterns (LBP) were additionally implemented for better representation. Another approach of MHI was executed in [58], where faces of each frame were isolated, and Local Binary Pattern (LBP), Local Phase Quantization (LPQ) and Edge Orientation Histogram (EOH) features were extracted. Then MHH was applied for capturing the dynamic motion of the features. In this work an extension of MHH was implemented, which is 1-D MHH that is computed on the feature vector sequence. Another approach of MHI is Landmark Motion History Images (LMHI) that was computed on sequences of 2D facial landmarks. In recent work [40] LMHI was utilized for depression assessment.

3.3 Gabor Filters and Facial Image Analysis

A linear filter such as Gabor is often utilized in various computer vision frameworks, such as face recognition [59] and facial emotion recognition [60] [61]. Such filter has also been applied for automatic facial expression analysis in depression [62] [63]. Maddage et al. [64] proposed an implementation that detects depressed and non-depressed subjects from video recordings. In this work "Gabor wavelet features extracted at the facial landmarks which were detected using landmark model matching algorithm". Cruz et al. [65] implemented Gabor energy filters, succeeding face characteristics detection such as mouth, eyes, and eyebrows. However, contours from face texture is also detected which are not significant to facial expressions. In order to eliminate contours from background texture anisotropic inhibition was implemented [66].

3.4 Appearance-Based Descriptors

This section includes the literature review of the implemented appearance-base descriptors which are Histogram of Oriented Gradients (HOG), Local Binary Patterns (LBP) and Local Phase Quantization (LPQ). Histogram of Oriented gradients (HOG) is an appearance-based descriptor that was initially used in image processing and computer vision framework for object detection [67] [68]. The idea behind HOG descriptor is that the object appearance and shape is described by the distribution of intensity gradients. The image is separated into small connected parts called cells. For each pixel of the cell, a histogram of gradient directions is compiled. The descriptor is the concatenation of these histograms. Concerning depression analysis, several implementations have been done [69], [58] using HOG.

Local Phase Quantization (LPQ) is a descriptor for texture classification. It was developed by Ojansivu et al. [70] to blur insensitive texture classification. Recently it was utilized successfully for facial expression recognition [71]. The idea behind LPQ is based on quantized phase of the discrete Fourier transform (DFT) computed in local image. The appearance-based descriptor has also been implemented for depression detection at the feature extraction step [72] [58]. An extension of LPQ is LPQ on Three Orthogonal Planes (LPQ-TOP) that considers patterns in three orthogonal planes: XY XT and YT and concatenates the pattern co-occurrences in these three directions.

In [73] LBP-TOP was suggested for automated depression detection. The original Local Binary Pattern algorithm (LBP) was introduced by Ojala et al. [74] and was proven a powerful mean for texture classification. Later, it was utilized for face detection [75] and for facial expression recognition [76]. An image is divided into small regions and each pixel's intensity is compared with those of its neighbors. A binary number is derived from this procedure and a histogram depending on the frequency occurring is produced. The LBP features can be fast derived as it is computational simple, whilst still retaining enough information. Thus, it is frequently used in recent bibliography for automated depression detection [77] [58]. LBP features have been used for detecting emotion expressions from the face. Jiang et al. [78] used LBP-based features to automatically recognize specific action unit. Several variants of LBP have also been proposed and implemented for depression assessment. Such a variant is LBP on Three Orthogonal Planes (LBP-TOP) that consider patterns in three orthogonal planes: XY XT and YT and concatenates the pattern co-occurrences in these three directions [77].

3.5 Available Datasets

In conducting a comprehensive study on depression assessment, obtaining a reference dataset one of the most significant steps, since such data are of utmost importance for algorithmic development. Such datasets include sensitive personal data and their publication usually involve ethical constraints. Researchers should be aware that it is always preferable to have explicit consent for processing such data. Due to the sensitive nature of the data, their open access is difficult or it is achieved under conditions. Usually research groups working in similar areas cannot openly share their data, which forces researchers of depression to collect their own dataset [79].

While conducting a data collection for depression analysis, depressed and non-depressed subjects should be carefully selected. The outmost importance criteria for depressed participants is that no other mental disorder history should coexist but only depression. As for control subjects, they should be chosen with no history of mental illness and to match their gender and their age with the depressed participants. It would be ideal to have a large clinical depression database available in order to develop a flexible and generalized, automatic depression detecting framework. However, this is not feasible and has not been achieved till now.

The Pittsburgh dataset was developed by the University of Pittsburgh and includes 57 interviews of both highly and mildly depressed patients, with audio and video recordings. The BlackDog dataset was introduced by the Black Dog clinical research institute focusing on mood disorders. The dataset has total 130 subjects (patients diagnosed with severe depression and healthy control) between 21 to 75 years old. Participants are recorded while reading sentences and giving an interview. The interview included specific question where the participants were asked to describe events that had aroused significant feelings.

The availability and the access to databases concerning depression is limited. Although there is a series of international Audio/Visual Emotion Recognition Challenges (AVEC 2013 [12], AVEC 2014 [13] and AVEC 2016 [14]) that are available after signing the license agreement. AVEC 2013 and AVEC 2014 are subset of the audio-video depressive language corpus (AViD-Corpus). AVEC 2013 and AVEC 2014 provide a dataset with vocal and visual raw data in contrast with AVEC 2016 (DAIC-WOZ[1]) that provides only landmarks. In accordance with the baseline paper [12], it is a dataset consisted by 340 video clips with 292 different subjects. There is only one person per video and every subject is performing a human-computer interaction task with Power Point guidance, while being recorded by a web camera and a microphone. Participants are undiagnosed volunteers, who had followed the collecting procedure. Data collection methods include vowel phonation, reading a speech, counting from 1 to 10, solving a task out loud, singing and narration of a personal story from the past and an imagined story stimulated by a picture [13]. Each recording has average length about 25 minutes. The participants' age is on average 31.5 and the age range is between 18 and 63 year old and were recorded one to four times.

DAIC-WOZ (Distress Analysis Interview Corpus - Wizard of Oz) is another available database consisted of multimodal semi-structured clinical interviews. Designed to support the diagnosis of psychological distress conditions such as, depression, anxiety, and post-traumatic stress disorder (PTSD) [80]. The dataset contains four different types of interviews, such as face-to-face, automated, teleconference and Wizard-of-Oz in each one differ the conducted procedure of the interview. Face-to-face type is an interview between the participant and the interviewer in the same room, in contrast to the teleconference session that the interviewer is over a teleconferencing system. The Wizard-of-Oz interviews, conducted by an animated virtual interviewer, was controlled by a human interviewer in another room. The automated interviews are conducted by a virtual interviewer in a fully automated mode. The participants had to answer PHQ-9 questionnaire [80] which is a diagnostic instrument for common mental disorders and monitor the severity of depression [81]. The questionnaires outcome reveal that 29% of the

[1] http://dcapswoz.ict.usc.edu/

participants are affected from depression. The dataset consists 189 sessions between 7 and 33 minutes.

3.6 Contribution of the present Thesis

The methodological framework proposed in the present thesis is evaluated on the AVEC dataset. Two motion representation methods were utilized with several combinations of appearance-based descriptors extracted from the face region. Relevant work [82] implemented on the same dataset, using SVM classifier and LGBP-TOP descriptor achieved 82% accuracy. In the proposed work we manage to highlight a comparable effectiveness score, which is 81.9%.

However, the main contribution of the present work is the proposed variant of Motion History Image, the Gabor Motion History Image which is a novel implementation for motion representation. Such an approach has never implemented before. Usually motion representation images are extracted from video images. Whereas in this work Gabor filters and the Gabor inhibited algorithm are applied on each video images.

4. Methodology

The implementation methodology followed during the development of the current thesis is included the chapter 4. The steps of the computational workflow involved are illustrated in Figure 3. The preprocessing is the initial step, which is followed by motion representation, which provides motion representation images. Those images are used for the feature extraction step. The last two steps are dimensionality reduction and classification. Additionally, this section includes details of the SW systems used for the implementation.

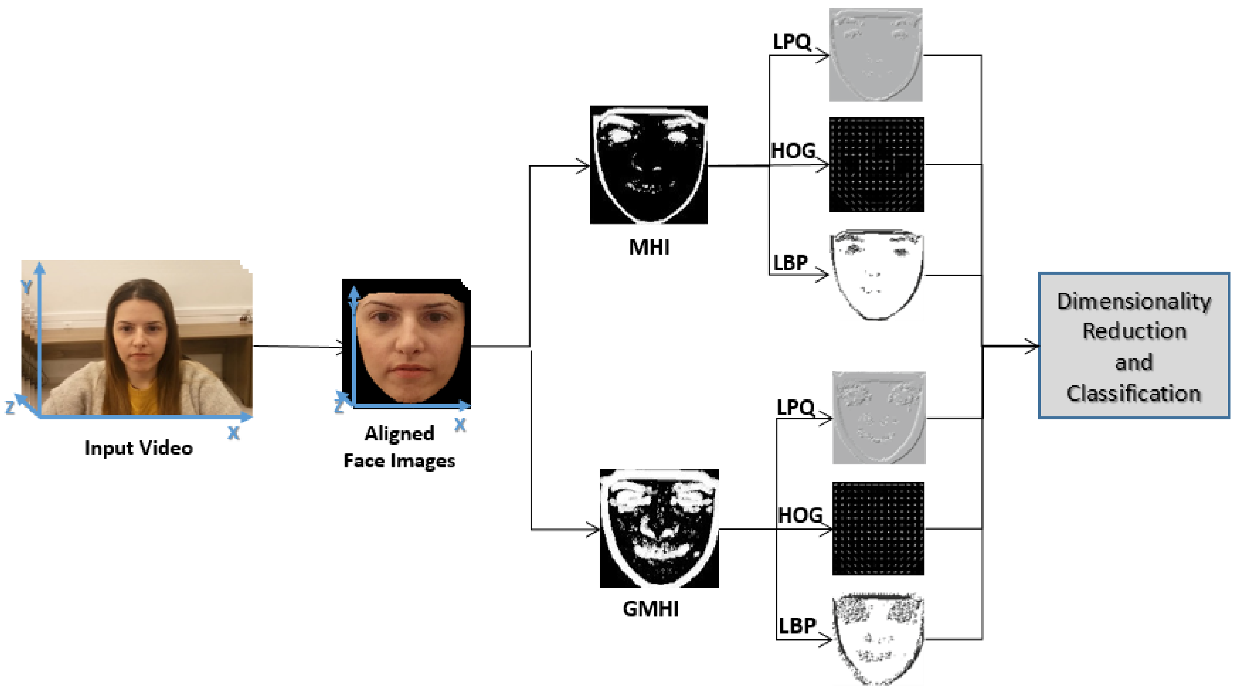


Figure 3 Pipeline of the Proposed Algorithm

4.1 Dataset Description

The Audiovisual Emotion Challenge (AVEC) dataset is a subset of the German audio-video depressive language corpus AVid that includes audio and video recordings [12]. The data has been collected for measuring and monitoring depression severity. In section 3.5 mentioned some general information about AVid and now is presented some details about the subset AVEC 2013.

The recordings where conducted using a laptop connected with a headset (sampling rate of 44, 100Hz, 16 bit) and a webcam. During the data collection procedure the participants were guided by a self-paced Power Point presentation to perform several tasks in German language, such as:

- Reading passage “Die Sonne und der Wind“
- Counting from 1 to 10
- Singing a German nursery rhyme
- Replying to a free-response question, telling their own story from the past. Some examples are “What is your favorite dish?”, “What was your best gift, and why?” and “Discuss a sad childhood memory”
- Speaking while solving a task
- Description of an imagined story applying the Thematic Apperception Test (TAT)
- Sustained vowel phonation, sustained loud vowel phonation, and sustained smiling vowel phonation

Depression severity is estimated as a continuous variable by using Beck Depression Inventory score (see section 2.1 Tools for Depression Assessment). A subject is classified in a category according to the depression severity level. The standard BDI cut-offs are:

- 0–13: Minimal depression
- 14–18: Mild depression
- 19–28: Moderate depression
- 30–63: Severe depression

The average BDI score in the AVEC dataset is 15 and the standard deviations is 12.3. Categorical assessment of video recordings is aiming to predict absence or presence of depression according to the BDI score. BDI score less than 13 points corresponds to non-depressed and more than 13 is depressed.

In the present work a subset of the dataset was used. It includes two tasks, the Northwind that includes the reading excerpt, and the Freeform that includes the questions’ answering. The recording sessions were split in three data partitions (training, development and test set), with 300 recordings totally. However in terms of the challenge annotations were made available only for 200 of the total recordings.

4.2 Preprocessing

This dissertation has been developed according to the proposed [51] standard pipeline for automated depression analysis which is illustrated in Figure 1. Once data have been collected, preprocessing is the first step of the method, where the Region of Interest (ROI) is detected. In our work this region of interest is the whole face. For the video preprocessing open source software OpenFace was used [83]. Using OpenFace aligned facial images sized 112 × 112 are extracted. OpenFace provides a binary success outcome for each aligned image, which specifies the successful and the unsuccessful facial detection. Score "1" represents successful facial landmarks detection, and "0" score the failure of facial landmarks detection. No more than the successfully detected aligned images are used for the present framework development. Figure 4 illustrates a video frame of size 1920×1080 and the extracted aligned face image sized 112×112. For illustration purposes in Figure 4 Preprocessing with OpenFaceFigure 4 both pictures have been resized.

Figure 4 Preprocessing with OpenFace

4.2.1 OpenFace

Baltrušaitis et al. [83] intended to bridge the gap between state-of-the-art algorithms and available toolkits. They presented an open source tool, for computer vision and machine learning researchers, giving the opportunity of building interactive applications based on facial behavior analysis. OpenFace utilizes facial action unit recognition, head pose estimation, facial landmark detection and eye-gaze estimation. It also gives the ability of performing all these tasks in real-time with a simple webcam and easy integration with other applications.

The proposed pipeline of OpenFace is illustrated in Figure 5 including facial landmark detection, head pose, eye gaze estimation, face extraction and facial action unit recognition. The core technology used for facial landmark detection is Conditional Local Neural Fields (CLNF) [84] which is an instance of a Constrained Local Model (CLM) [85].

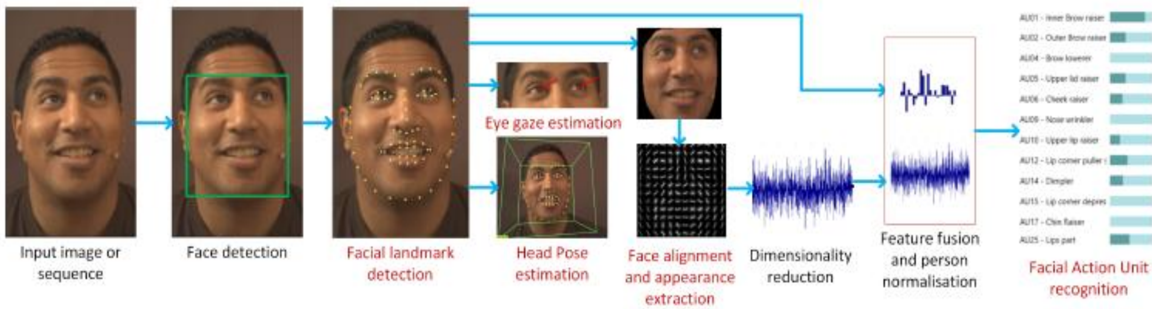


Figure 5 OpenFace facial behavior analysis pipeline [83]

4.3 Motion Representation

In the present work the implemented methods derived from two different motion representation algorithms. (1) Motion History Image (MHI) and (2) Gabor Motion History Image (GMHI). Figure 6 illustrates the Motion History Image Representation and Figure 9 represents Gabor Motion History Image.

4.3.1 Motion History Image

In the present work a view-based approach is developed for motion representation

"*The Motion History Image is explained and presented as a concept by which moving parts of a video sequence can be engraved in a single image, from where one can predict the motion flow as well as the moving parts of the video action*" [55].

It is a static, scalar-valued image where intensity is a function of recency of motion [86]. Motion is condensed into gray scale images in which the most recent action in a video sequence is illustrated with white pixels. Gray scalar values represent movement that happened less recently.

The MHI algorithm is applied at the aligned face image sequence that were extracted in the prepossessing step (section 4.2) using the OpenFace toolkit. The construction of the Motion History Image follows:

Equations (1) and (2) describe two consecutive aligned images

$$I(x,y,t) - b_t(x,y) + m_t(x,y) + n_t(x,y) \qquad (1)$$

$$I(x,y,t+1) = b_{t+1}(x,y) + m_{t+1}(x,y) + n_{t+1}(x,y) \qquad (2)$$

Where,

$b_t(x, y)$: Static background for t_{th} frame

$m_t(x, y)$: Moving objects for t_{th} frame

$n_t(x, y)$: Background noise for t_{th} frame

Now, if we consider consecutive frame differencing approach for extracting moving objects, we can get,

$$diff(x, y, t) = I(x, y, t + 1) - I(x, y, t)$$
$$diff(x, y, t) = b(x, y) + md(x, y, t) + nd(x, y, t)$$

Where,

$b(x, y, t)$: Overlapped area in consecutive frames

$md(x, y, t)$: Motion region

$nd(x, y, t)$: Noise

The MHI can be generated using difference of frames (DOF),

$$\Psi(x, y, t) = \left\{ \begin{matrix} 1 & if\ D(x, y, t) > \xi \\ 0 & otherwise \end{matrix} \right\}$$

Where,

$\Psi(x, y, t)$: Is the binarization of the difference of frames by considering a threshold ξ

ξ: Is the minimal intensity difference between two images

$D(x, y, t)$: Is defined with difference distance Δ as,

$$D(x, y, t) = |I(x, y, t) - I(x, y, t \pm \Delta)|$$

The MHI $H_T(x, y, t)$ is computed:

$$H_T(x, y, t) = \left\{ \begin{matrix} \tau & if\ \psi = 1 \\ max(0, H_T(x, y, t - 1) - 1) & otherwise \end{matrix} \right\}$$

Where,

(x, y), t: Is the pixel position and the time

τ : Defines the temporal duration of the MHI.

Motion History Image method was implemented by empirically setting the value ξ to 25. This configuration was set so that the static background is not represented in the MHI. Figure 6 depicts an implemented example of the developed MHI method.

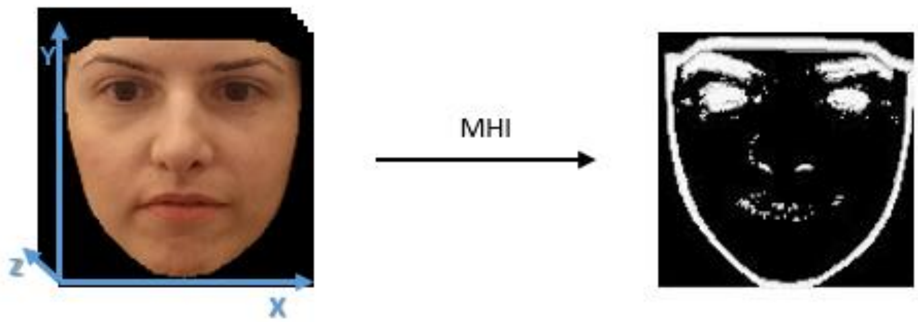


Figure 6 Motion History Image Representation

4.3.2 Gabor Motion History Image

The second motion representation algorithm implemented in the present work is Gabor Motion History Image. The algorithm has the same operation as the described MHI in section 4.3.1. For the GMHI computation Gabor inhibited images were utilized as they depicts the most relevant and important image information reducing the effect of background texture. The GMHI algorithm is applied at the aligned face image sequence of each video that were extracted in the prepossessing step (section 4.2). The construction of Gabor Motion History Image is described below.

In each aligned image a bank of Gabor filters with multiple orientations and wavelengths are applied.

$$g_{\lambda,\theta,\varphi,\sigma,\gamma}(x,y) = e^{-\frac{\tilde{x}^2+\gamma^2\tilde{y}^2}{2\sigma^2}} \cos\left(2\pi\frac{\tilde{x}}{\lambda}+\varphi\right)$$

With,

$$\tilde{x} = x\cos\theta + y\sin\theta$$
$$\tilde{y} = -x\sin\theta + y\cos\theta$$

Where,

x, y: Pixel location

λ: Wavelength

θ: Orientation

φ: Phase

σ: Standard deviation of the Gaussian

γ: Spatial aspect ratio

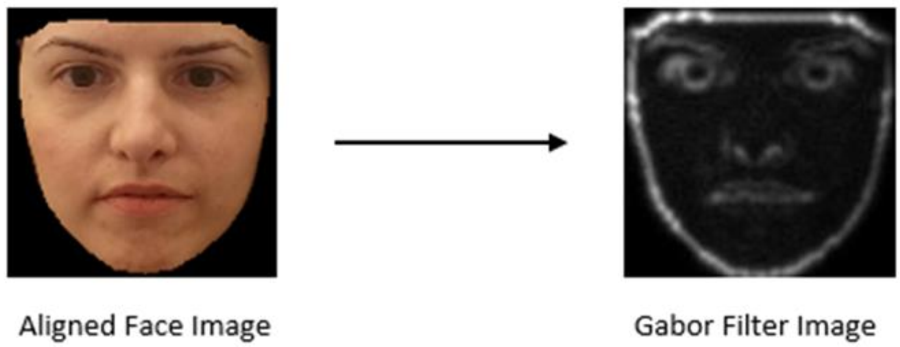


Figure 7 Gabor Filter Image

An example of applying Gabor filter on an aligned face image is illustrated in Figure 7, where contours of the face are measured. However, not all contours are significant to facial expression. In order to reduce the effect of background texture, anisotropic inhibition was developed [66].

Background texture is estimated and then it is subtracted from the Gabor filter image

$$w(x, y) = \frac{1}{\|g(DoG)\|_1} g\big(DoG(x, y)\big)$$

Figure 8 Gabor Inhibited Image

Figure 8 depicts the above mentioned procedure. The original aligned facial image is processed by the implementation of Gabor filters and of the anisotropic inhibition. It is clearly illustrated that Gabor filters has successfully detected face contours (mouth, eyes and eyebrows), while anisotropic inhibition removed the background texture.

Gabor Motion History Image method was developed defining empirically threshold to 8. This configuration was set so that the static background is not represented in the GMHI. Applying

Gabor filters and the Gabor inhibited algorithm to the aligned face image of a video compute motion representation GMHI. An example is illustrated in Figure 9.

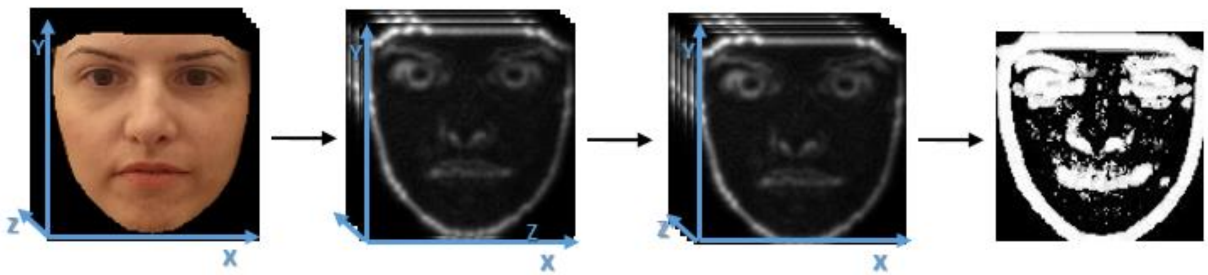


Figure 9 Process flow of Gabor Motion History Image

4.4 Feature Extraction

One of the main issues in designing automatic depression detection systems is feature selection that can achieve the desired effectiveness. In this section the selected descriptors utilized for feature extraction are analyzed. The Local Binary Patterns (LBP), the Local Phase Quantization (LPQ) and the Histogram of Oriented Gradients (HOG) are the implemented appearance-based descriptors in the present work. Some statistical features have been also extracted, such as mean, standard deviation and histogram.

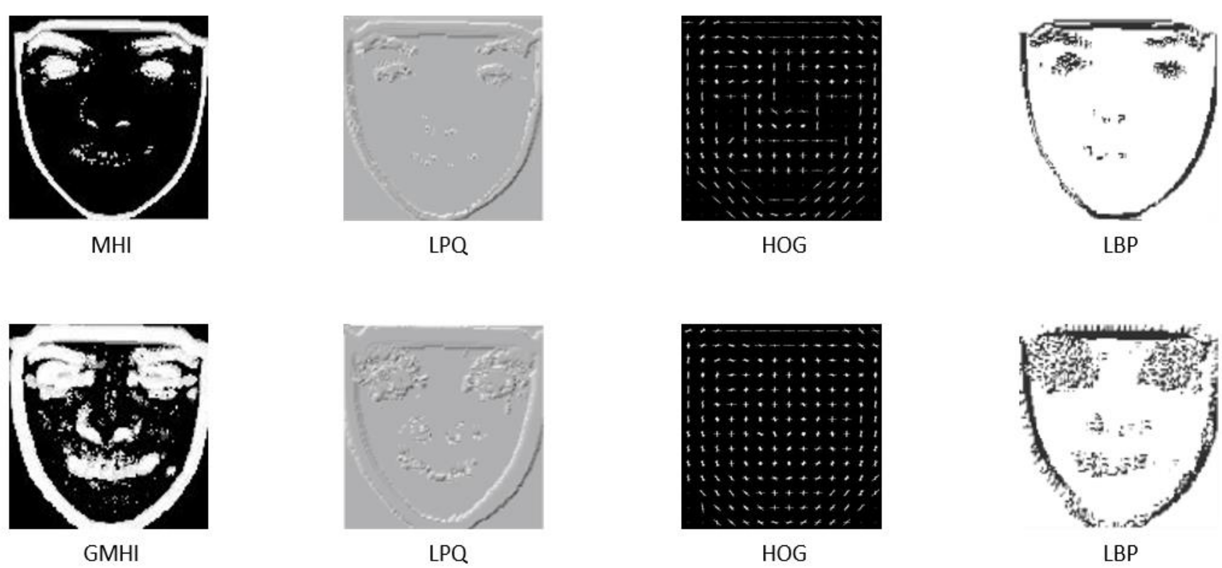


Figure 10 Visualization of appearance descriptors' responses

4.4.1 Histogram of Oriented Gradients

HOG stands for Histograms of Oriented Gradients is an appearance feature descriptor proposed by Dalal et al. [87]. This method was originally employed for human detection in static images, and more widely on object detection. The main algorithm is based on edge information [67] that is described by the distribution of intensity gradients or edge directions. The source picture is divided into a dense grid of small region called cells (8×8 pixels). For each pixel within the cell the vertical and horizontal gradients are computed using 1-D Sobel vertical and horizontal operators.

$$[-1,0,1] \; and \; [-1,0,1]^T$$

Cells are described by the local distribution of the edge orientations and the corresponding gradient magnitude. The cells can be either rectangular or radial and the histogram spread over 0 to 360 degrees. A local histogram of oriented gradients is formed, where the magnitude of the edge gradient for each orientation bin is computed for all pixels. The extracted HOG feature implemented on MHI and GMHI results a 1×6084 feature vector (Figure 11).In Figure 12 a visualization of HOG features extracted from MHI and GMHI is illustrated.

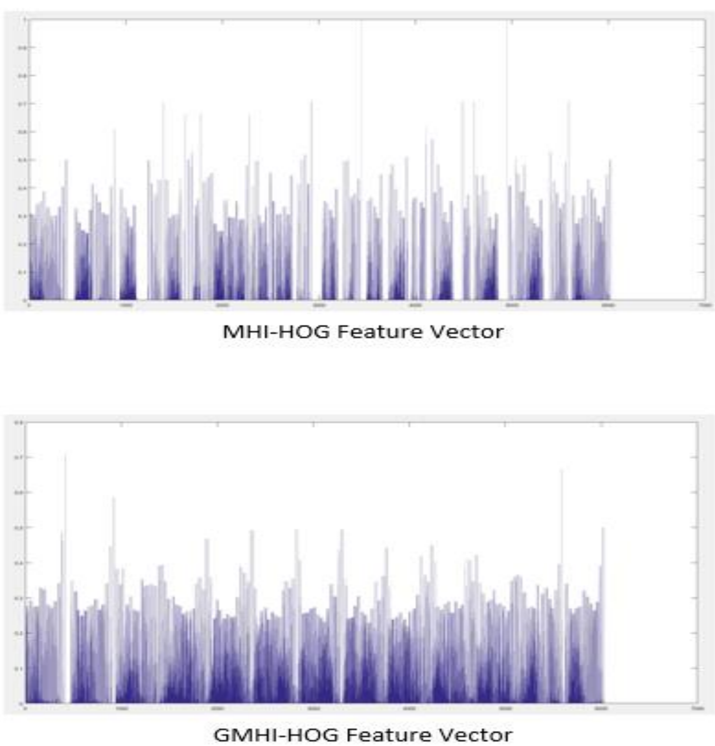


Figure 11 Plot of HOG Feature Vector for MHI and GMHI

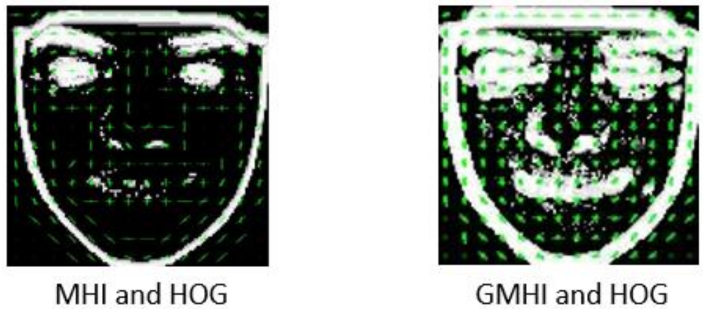


Figure 12 Visualization of HOG descriptor for MHI and GMHI

4.4.2 Local Binary Patterns

Ojala et al. [74] originally proposed the Local Binary Patterns descriptor. It is an efficient texture operator which aims to encode the local structures around each pixel. For each pixel a binary vector is computed by the comparison of the pixel's intensity with those of its neighbors. The image is divided into overlapping cells (neighborhoods) usually consisted by 3×3 pixels. The center pixel value is subtracted with its eight neighbor's value following the pixels along a circle (clockwise). Where the center pixel's value is greater than the neighbor's value, the pixel is encoded to "0", otherwise to "1". This procedure results in an 8-digit binary number that is converted to decimal. An example of the basic LPB operation is depicted in Figure 13.

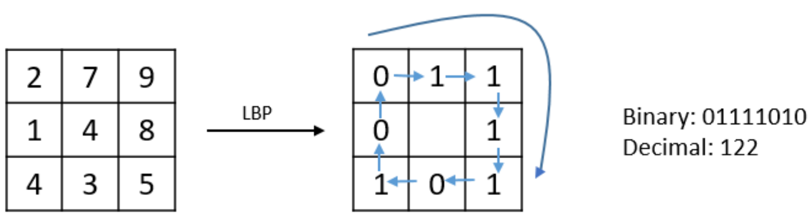


Figure 13 An Example of the LBP Operation

A histogram of the frequency of each decimal number occurring is calculated. Some binary patterns are more common in texture images than other. A binary pattern that contains at most two 1-0 or 0-1 transitions is called uniformed, otherwise it is called non-uniform. For the histogram computation the uniform pattern has a separate bin and all non-uniform patterns are assigned to a single bin [88].

The LBP operator was extended to use neighborhoods of different sizes. It was achieved giving the opportunity of using any radius and number of pixels in the neighborhood. The LBP notation is (P, R) where P is used for pixel neighborhood and R is for radius [89]. LBP can be expressed as:

$$LBP_{P,R}(x_c, y_c) = \sum_{P=0}^{P-1} s(i_p - i_c)2^p$$

Where,

(x_c, y_c): Pixel's location

i_c and $i_p$are values of the central pixel

P: Surrounding pixels in the circle neighborhood

R: Radius

$$s(x) = \begin{Bmatrix} 1 & if & x \geq 0 \\ 0 & if & x < 0 \end{Bmatrix}$$

Two set of LBP descriptor was implemented in the present work. The first set is radius=1 and neighborhood=8, where the total patterns are 256, 58 of which are uniform, which yields in 59 different bins ().The second is radius=2 and neighborhood=16 with 243 bins.

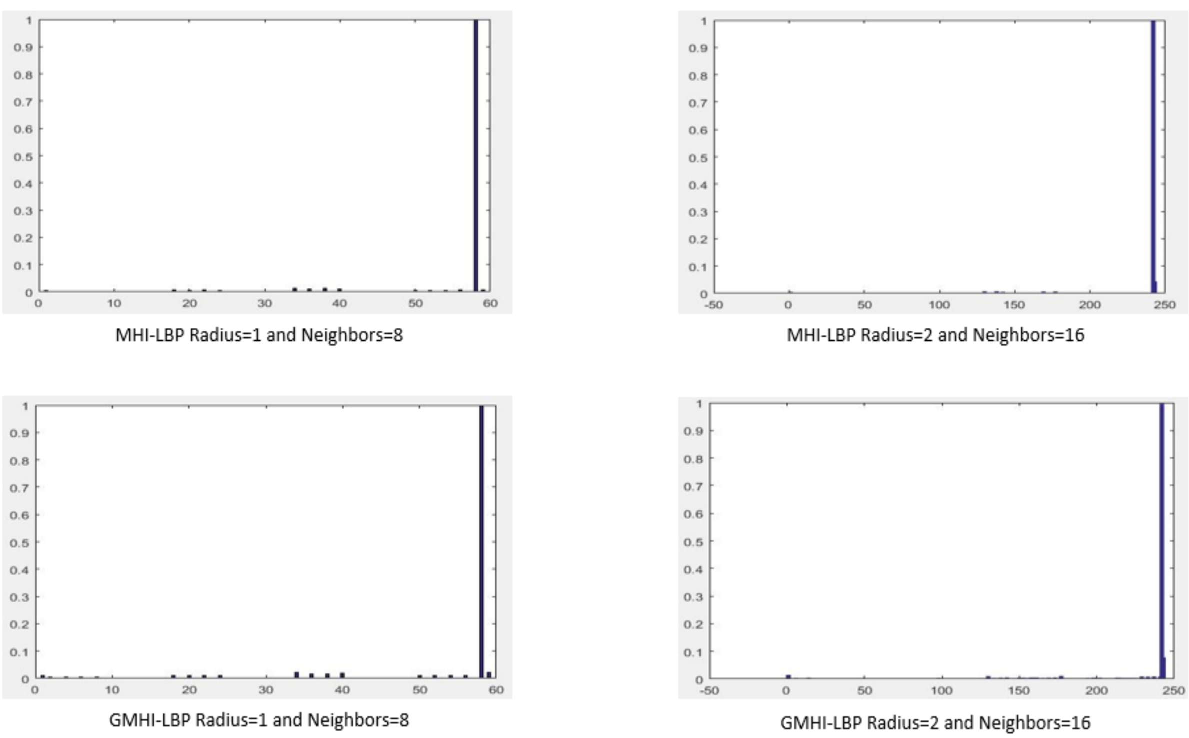


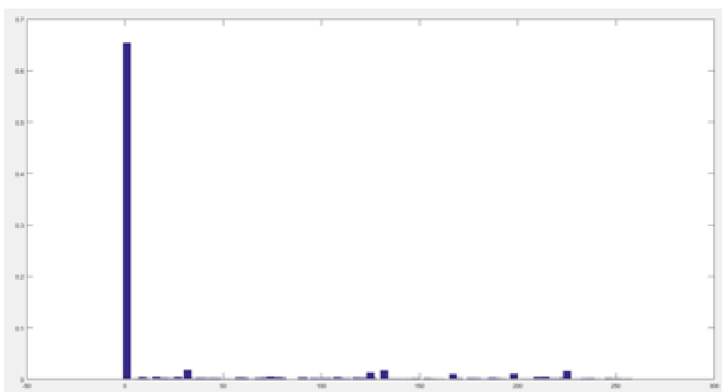
Figure 14 LBP Histograms for MHI and GMHI

4.4.3 Local Phase Quantization

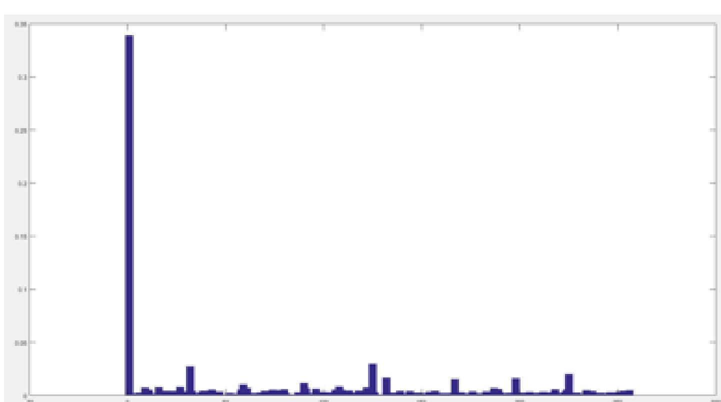
The Local Phase Quantization (LPQ) is an appearance-based descriptor proposed by Ojansivu et al. [70]. It was initially designed for blur insensitive texture classification. However recently it was successfully applied for facial expression recognition [71]. LPQ is based on the blur invariance property of the Fourier phase spectrum. The image is divided into blocks where a short-time Fourier transform (STFT) is applied to extract local phase information.

$$F(u,x) = \sum_{y \in N_x} f(x-y)\, e^{-j2\pi u^T y}$$

The Fourier transform is computed for the four sets of coefficients and then quantization based on the sign of the coefficient follows. The face image is divided into sub-regions for which individual 256-dimensional histograms are computed by binning the LPQ coefficients. Figure 15 Illustrates the LPQ feature extraction for MHI and GMHI.



MHI-LPQ



GMHI-LPQ

Figure 15 LPQ Feature extraction from MHI and GMHI

4.4.4 Statistical Features

It was not only appearance-based descriptors that were implemented, but also same statistical features were utilized. Mean (average value), standard deviation (mean deviation of the signal compared to the average) and combined histogram are extracted from motion representation images (MHI and GMHI) and regarded as a single descriptor [HIST-MEAN-STD. (1) and (2) are the equations for mean and standard deviation respectively. Histogram is based on the picture value intensities. Each bin represents a different value intensity, zero values are overlooked and bins are computed with 255 gray values.

$$Mean = \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right) \qquad (1)$$

$$Standard\ Deviation = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N-1}} \qquad (2)$$

4.5 Feature Selection

Feature extraction algorithms usually generate long features vectors with high dimensionality power. Feature selection is the process of choosing the optimum subset of features for classification. Dimensionality reduction methods are applied to the extracted features so as to use the highly discriminating features and omitting the noncontributing features [79]. In order to reduce the computational time of classification and the irrelevant features we use a Principal Component Analysis (PCA) model, where new variables are linear combinations of the original feature, chosen to capture as much on the original variance as possible [90]. For a feature matrix m×n, with m rows (samples) and n columns (features), the principal components vectors are the eigenvectors of the n×n coefficient matrix, ordered by decreasing magnitude of the corresponding eigenvalue. PCA method reduces feature dimensionality of the original data by projecting the data into a lower dimensional space. The smaller feature set is used as input for the classification step.

4.6 Classification

In the classification step, the feature vector that is provided by the feature extraction step is used by a selected classifier in order to evaluate the proposed system, and to assign the unclassified object a category according to the established criteria. When trying to establish a relationship between features mistakes can be made. Machine learning can improve the efficiency of such systems [91]. Machine learning is a computer science field that enables computers to predict an outcome just applying a set of algorithms. In [91] the difference between supervised and unsupervised learning is simply highlighted. If instances are given with known labels (the corresponding correct outputs) it is supervised learning. On the other hand unsupervised learning instances are unlabeled.

In supervised learning the input data for training the model are labeled and have a corresponding relevance to the testing data. The algorithm creates a model according to the training set and can make predictions for any unknown given data. This method can be used to generalize from new instances. The employed supervised learning classifiers are Support Vector Machines (SVM), Naïve Bayes, Nearest Neighbors (kNN), and Random Forest.

k-Nearest Neighbors (kNN)

The k-Nearest Neighbors algorithm is a non-parametric algorithm used for classification and regression. This method (Figure 16) uses distance metrics (i.e Euclidean, Cityblock, Chebychev etc.) for finding the nearest samples in the feature space. The value k defines the number of classes included in the voting scheme. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors.

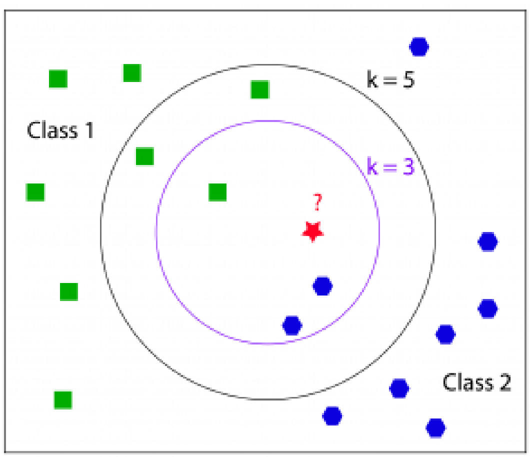


Figure 16 k-Nearest Neighbors Overview [113]

For SVM classification in Matlab the following configuration parameters[2] were used:

- Distance Metric = Seuclidean (Standardized Euclidean distance)

[2] https://www.mathworks.com/help/stats/fitcknn.html#d119e294908

- Number of Neighbors =6

Random Forest

Random forest is a method for classification and regression. Its operation (Figure 17) relies on developing a multitude of decision trees by using randomly selected training data and randomly selected variables. Each tree a unit vote is cast for the most popular class to classify an input vector [92].

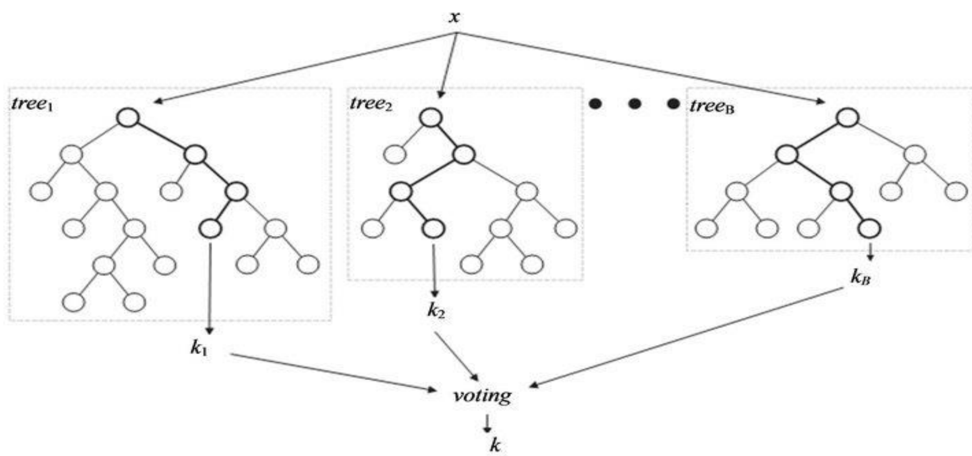


Figure 17 Random Forest Overview

Support Vector Machine (SVM)

SVM classifier is a supervised learning model for binary classification. The algorithm constructs an optimal hyper-plane or set of hyper-planes that separates the data points of one feature class from those of the other feature class. The objective of SVM modelling is to find the optimal hyper plane that separates clusters providing the highest margin (distance between the separating hyper-plane) distance between the nearest points of the classes [93].

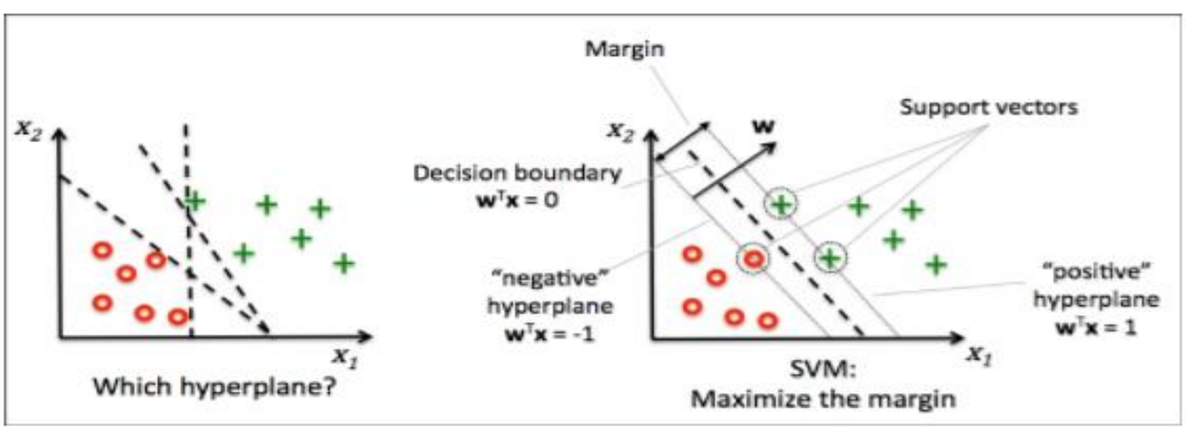


Figure 18 SVM Classifier Overview [110]

For SVM classification in Matlab the following configuration parameters3 were used:

- Kernel Function = Gaussian
- Outlier Fraction = 10
- Score Transform=$log\left(\frac{x}{(1-x)}\right)$

Naïve Bayes

The Naïve Bayes is probabilistic technique based on the so-called Bayesian theorem with strong (naïve) independence assumption between the features. It is particularly suited for high input dimensionality. This method (Figure 19) assumes that the effect of an attribute value on a given class [92]. The Naïve Bayes algorithm can be used for classification and regression.

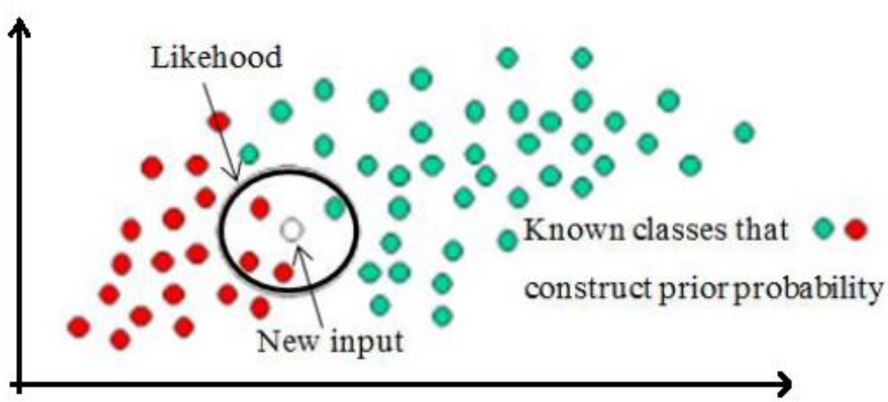


Figure 19 Naive Bayes Overview

[3] https://www.mathworks.com/help/stats/fitcsvm.html

5. Experimental Results

5.1 Cross Validation Method

Evaluation in depression recognition determines the quality of the system and identifies system's weaknesses. Cross-validation is statistical method of evaluation. The basic form of cross validation is k-fold cross validation where data are separated in k equal-sized sets (folds). A fold is secluded and the remaining are used for classification training. The removed data are used for performance evaluation. This process is repeated until all folds are tested [94]. Leave-One-Out (LOO) cross-validation method is a special case of k-fold cross validation, where k value is the total number of the data samples. A single instance from the original dataset is used for testing set and the remaining constitute the training set.

In depression recognition using LOO cross-validation might be unreliable as training set may include instances of the testing set, if the dataset includes several observations of the same subject. In order to avoid this issue, Leave One Subject Out cross validation method is used. Such method is person independent and each subject instance of the testing set is excluded from the training set. This method is utilized in the present work, as there are subjects with several recordings in the AVEC dataset.

5.2 Evaluation Measures

The evaluation of the system performance can be assessed with the use of confusion matrix [95]. In predictive analytics, a table of confusion, also called a confusion matrix, is a table with two rows and two columns that reports the number of false positives, false negatives, true positives, and true negatives. This allows more detailed analysis than mere proportion of correct classifications (accuracy) (Table 1).

True Positive (TP): TP: The predicted outcome is correct, and the instances are classified as positive.

False Negative (FN): FN: The predicted outcome is wrong, and the instances are classified as negative while they are positive.

False Positive (FP): The predicted outcome is right, and the instances are classified as positive while they are negative.

True Negative (TN): The predicted outcome is not in the actual class where the actual is not in than class.

	Predicted Class	
Actual Class	True Positive (TP)	False Negative(FN)
	False Positive (FP)	True Negative (TN)

Table 1 Confusion Matrix

Based on the confusion matrix, several statistical measures could be computed, such as accuracy, precision, sensitivity, specificity and F1 score.

Accuracy value defines the number of instances that have been correctly classified.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision value defines the degree of positive predicted positive values.

$$Precision = \frac{TP}{TP + FP}$$

Sensitivity or recall defines (TP rate) the proportion of positives that are correctly identified as positives.

$$sensitivity = recall = \frac{TP}{TP + FN}$$

Specificity (TN rate) defines the proportion of negatives that are correctly identified as negatives.

$$Specificity = \frac{TN}{TN + FP}$$

F-score or F1 score defines the harmonic mean of precision and sensitivity.

$$F1 - score = F - score = 2 \times \frac{Precision\ \times Recall}{Precision + Recall}$$

5.3 Classification and Evaluation

This subsection summarizes the performance of the proposed framework. Several configuration set of parameters were implemented trying to optimize the overall results, succeeding the maximum F1 score. We end up using four different classifiers which are k Nearest Neighbor, Random Forest, Support Vector Machine, and Naïve Bayes. Leave-One-Subject-Out cross validation method was used as more than one recording for each subject exist. The extracted features are HOG, LBP{1,8}, LBP{2,16}, LPQ and HIST-MEAN-STD, that produce 31 different combinations. Those combinations are used as feature vector input. The selected values for Principal Component Analysis are 10, 50, 100 and 150.

The following table (Table 2) summarizes the selected parameters for each classifier.

Classifier	Parameters
K Nearest Neighbor	Distancetypes=[{'seuclidean'}], Numberofneighbours= [6]
Random Forest	Default
SVM	KernelFunction = [{'gaussian'}], OutlierFraction = [0.1], ScoreTransform = [{'invlogit'}]
Naïve Bayes	Default

Table 2 Set of Parameters for Each Classifier

k-Nearest Neighbors (kNN) Classifier

Classifier k-Nearest Neighbors was employed for both MHI and GMHI motion representation for PCA 10, 50,100 and 150. Figure 20 and Figure 21 depicts the results of F1 score for each motion representation method. As for MHI, maximum F1 score 54.19% was achieved with PCA 50 for feature vectors HOG-HIST-MEAN-STD, LBP{1,8}-HOG-HIST-MEAN-STD, LBP{1,8}-LBP{2,16}-HOG-HIST-MEAN-STD, LBP{1,8}-LBP{2,16}-LPQ-HOG-HIST-MEAN-STD, LBP{1,8}-LPQ-HOG-HIST-MEAN-STD, LBP{2,16}-HOG-HIST-MEAN-STD, LBP{2,16}-LPQ-HOG-HIST-MEAN-STD and LPQ-HOG-HIST-MEAN-STD. For GMHI maximum F1 score is 58.46% with PCA 50 for feature vector LBP{1,8}-LBP{2,16}-LPQ-HIST-MEAN-STD and LBP{2,16}-LPQ-HIST-MEAN-STD.

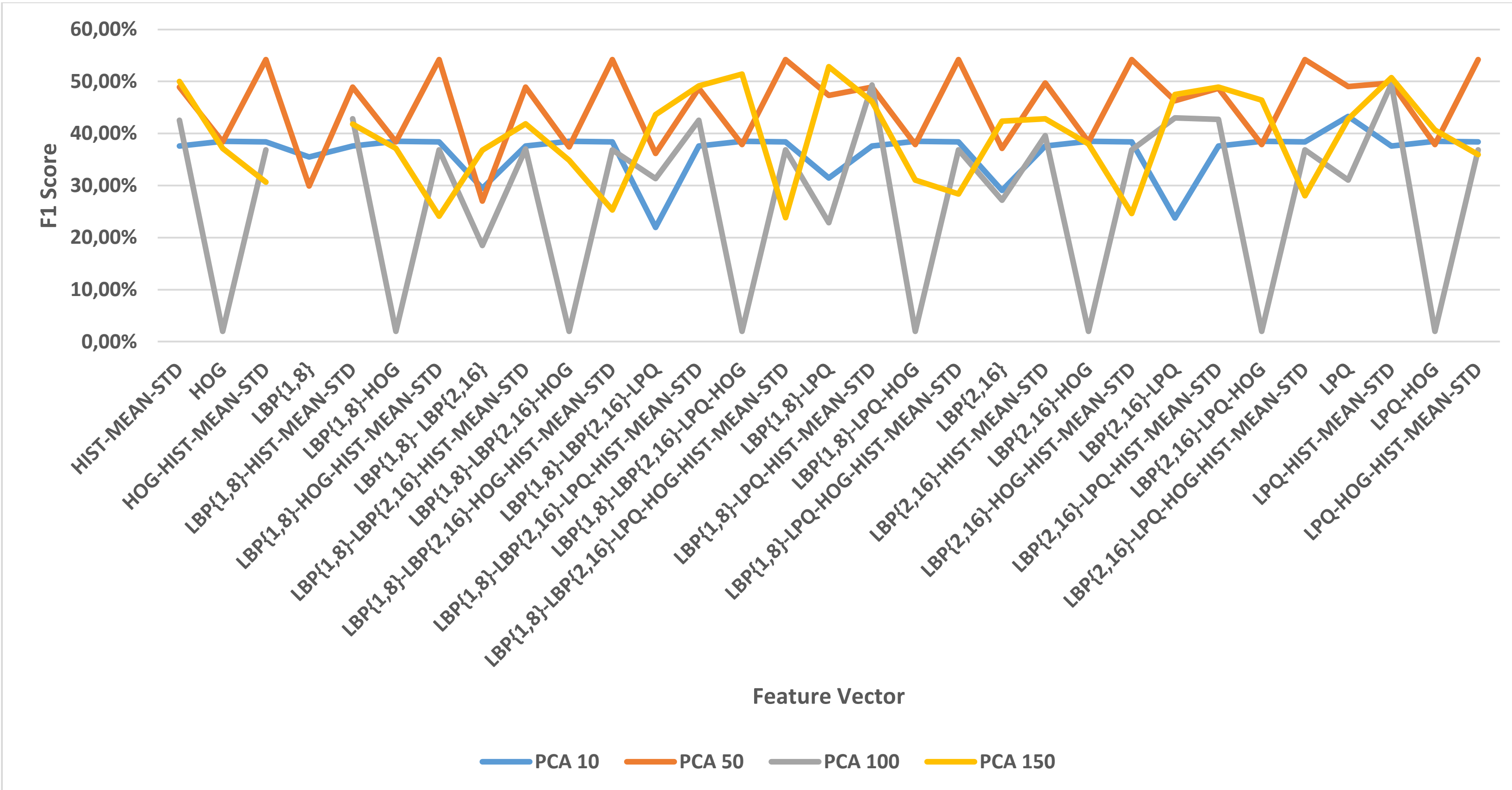


Figure 20 F1 Score Results for MHI with k-Nearest Neighbors Classifier

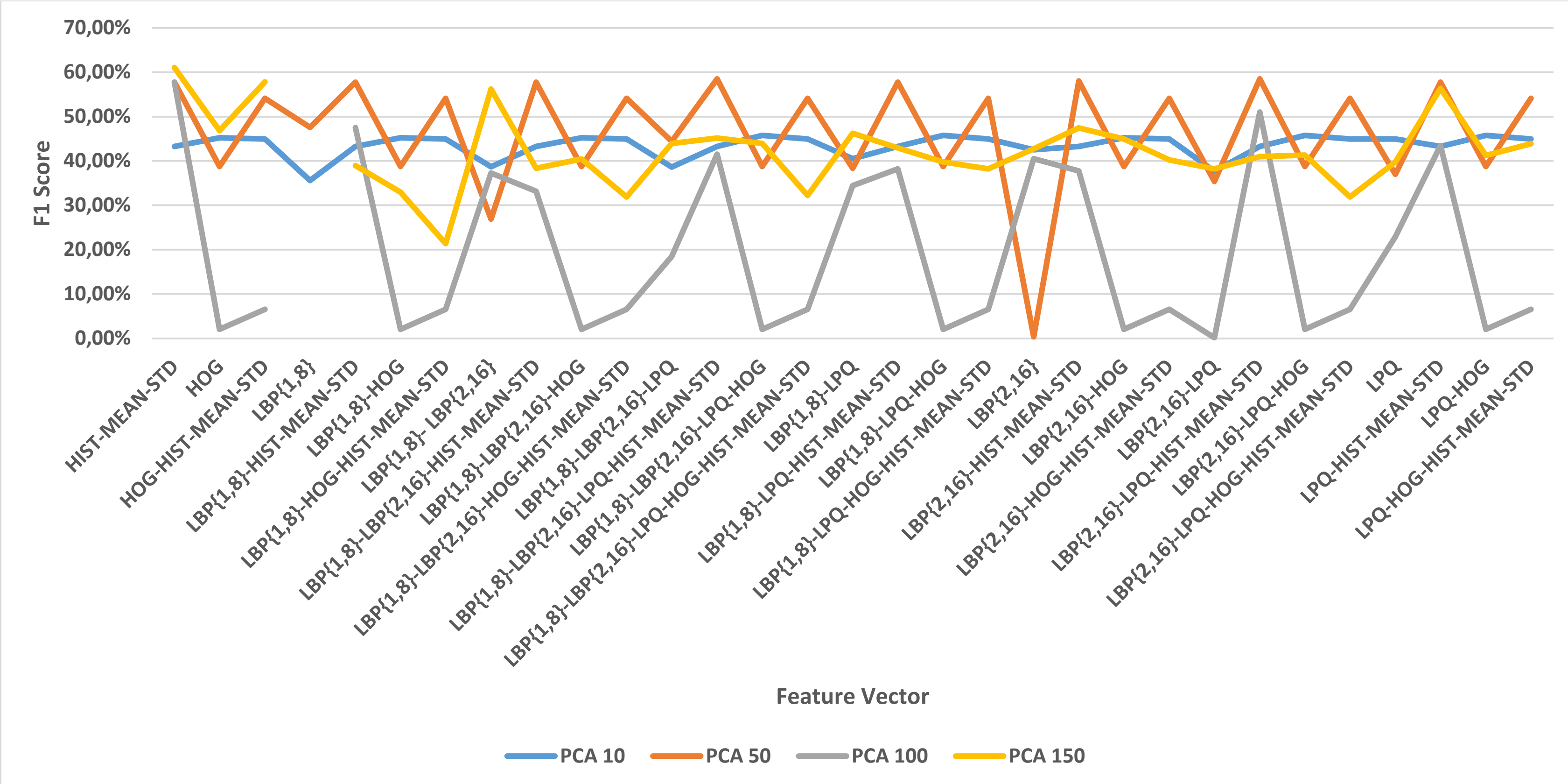


Figure 21 F1 Score Results for GMHI with k-Nearest Neighbors Classifier

Random Forest

Classifier Random Forest was employed for both MHI and GMHI motion representation for PCA 10, 50,100 and 150. Figure 22 and Figure 23 depicts the results of F1 score for each motion representation method. As for MHI, maximum F1 score 61.22% was achieved with PCA 50 for feature vectors LBP{1,8}-LBP{2,16}-LPQ-HOG-HIST-MEAN-STD and LBP{2,16}-LPQ-HOG-HIST-MEAN-STD. GMHI maximum F1 score is 62.75% with PCA 150 for feature vector LBP{2,16}-HIST-MEAN-STD.

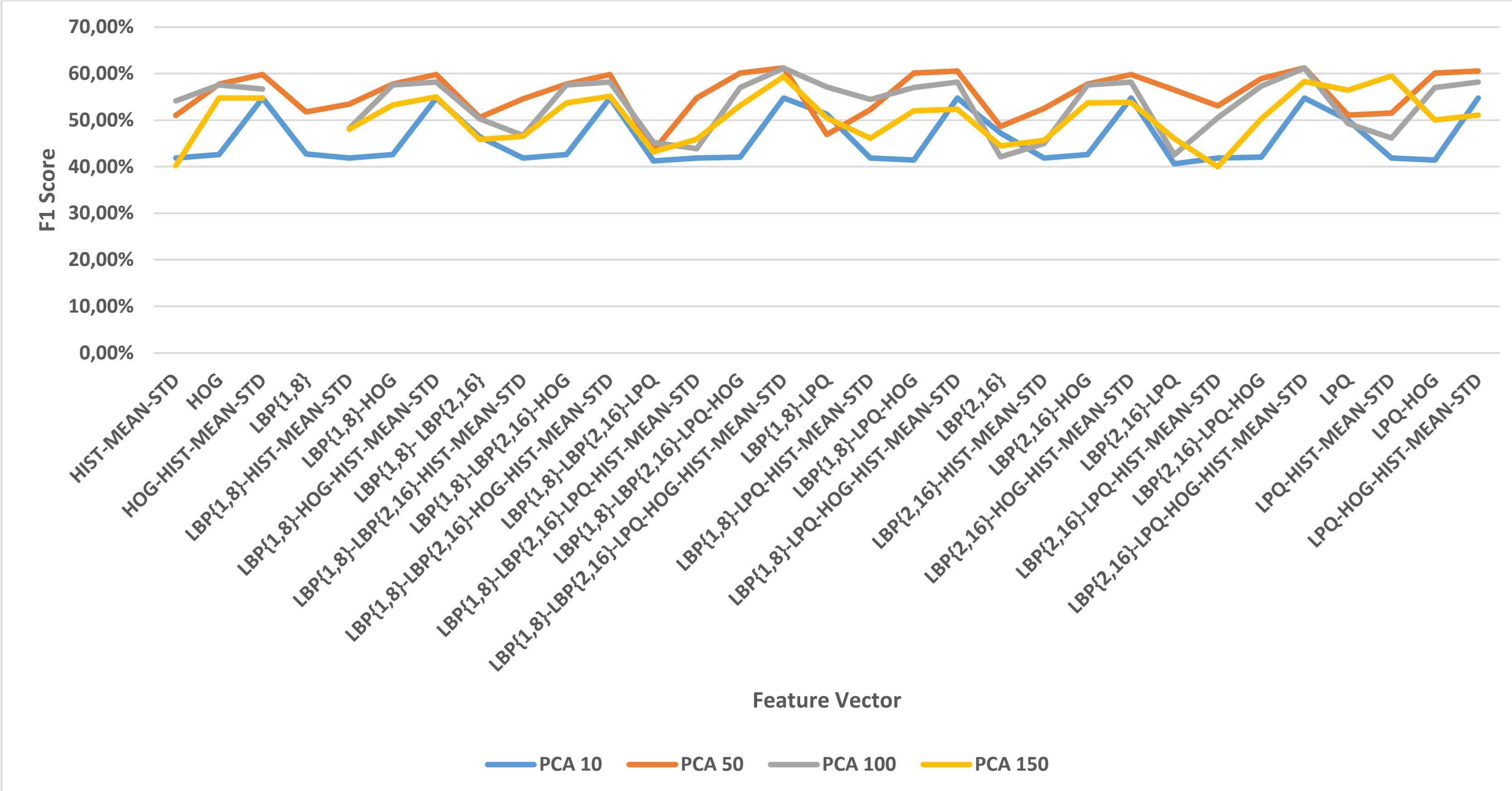


Figure 22 F1 Score Results for MHI with Random Forest Classifier

Figure 23 F1 Score for GMHI with Random Forest Classifier

Support Vector Machine (SVM)

Classifier Support Vector Machine was employed for both MHI and GMHI motion representation for PCA 10, 50,100 and 150. Figure 24 and Figure 25 depicts the results of F1 score for each motion representation method. As for MHI, maximum F1 score 81.93% was achieved with PCA 100 for feature vectors HIST-MEAN-STD and HOG. GMHI maximum F1 score is 81.93% with PCA 10 for feature vector HIST-MEAN-STD.

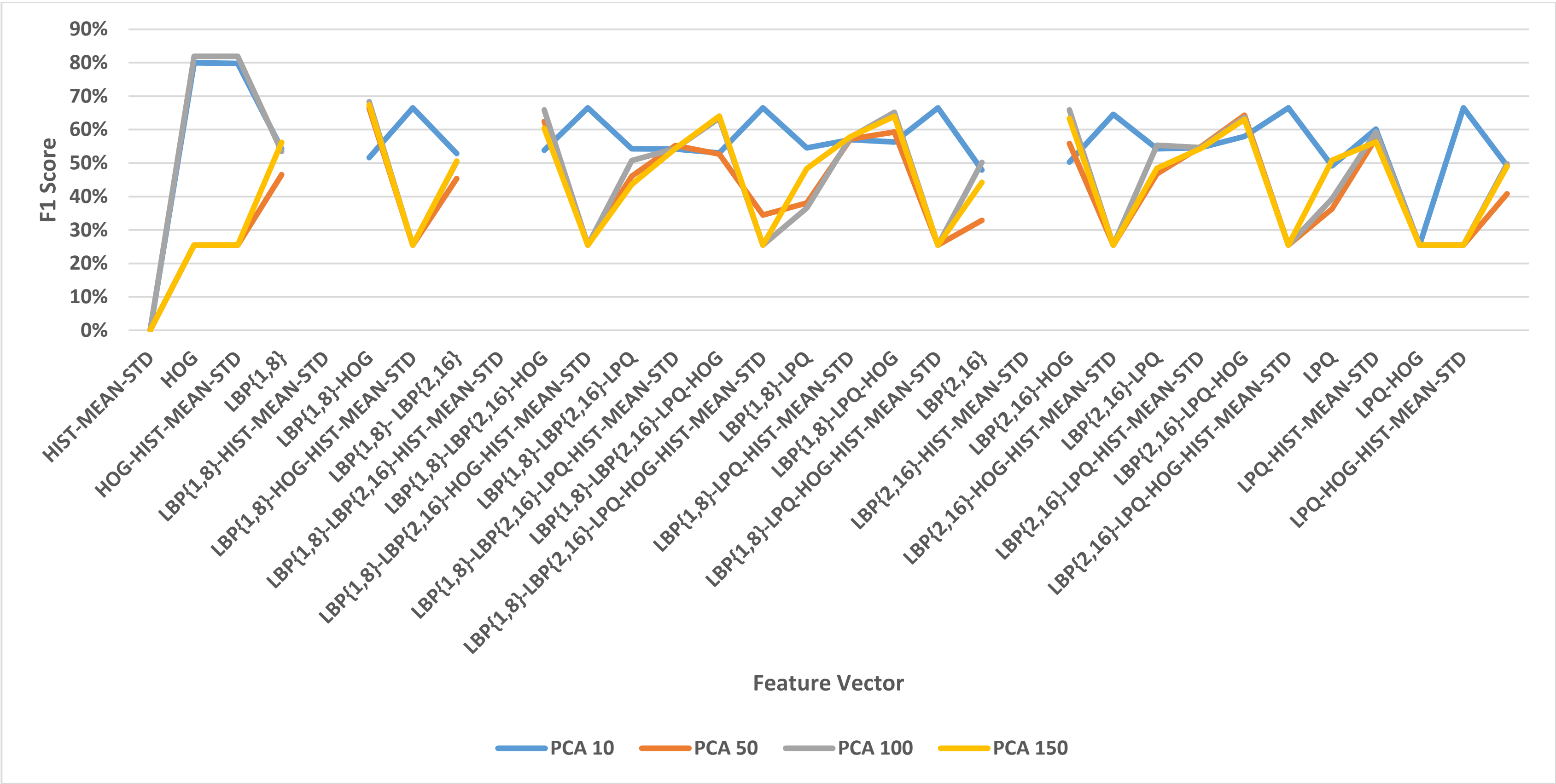


Figure 24 F1 Score for MHI with Support Vector Machine Classifier

Figure 25 F1 Score for GMHI with Support Vector Machine Classifier

Naïve Bayes

Classifier Naïve Bayes was employed for both MHI and GMHI motion representation for PCA 10, 50,100 and 150. Figure 26 and Figure 27 depict the results of F1 score for each motion representation method. As for MHI, maximum F1 score 66.19% was achieved with PCA 100 for feature vectors LBP{1,8}-LBP{2,16}. GMHI maximum F1 score is 66.67 % with PCA 150 for feature vector HOG-HIST-MEAN-STD.

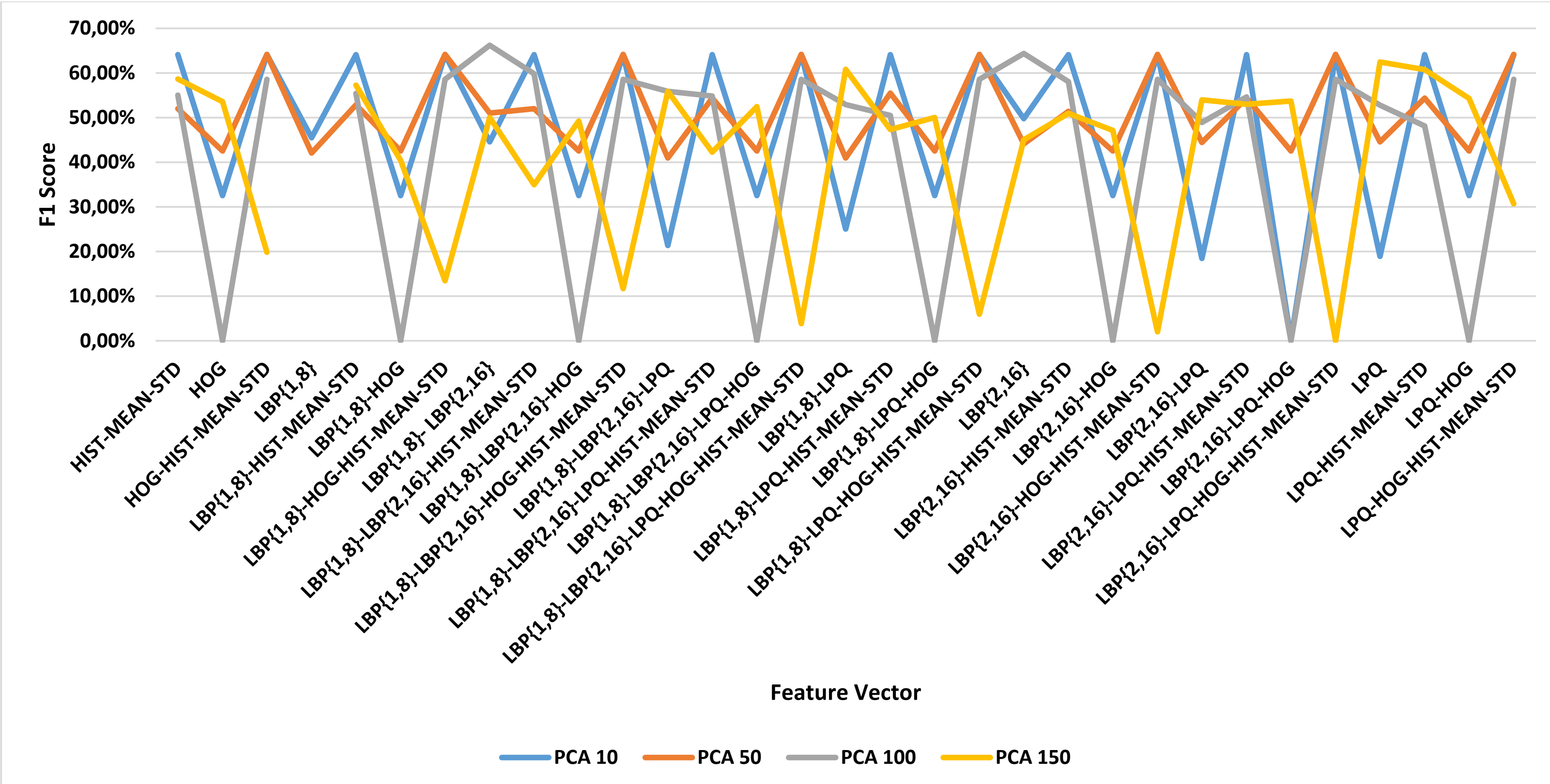


Figure 26 F1 Score for MHI with Naive Bayes Classifier

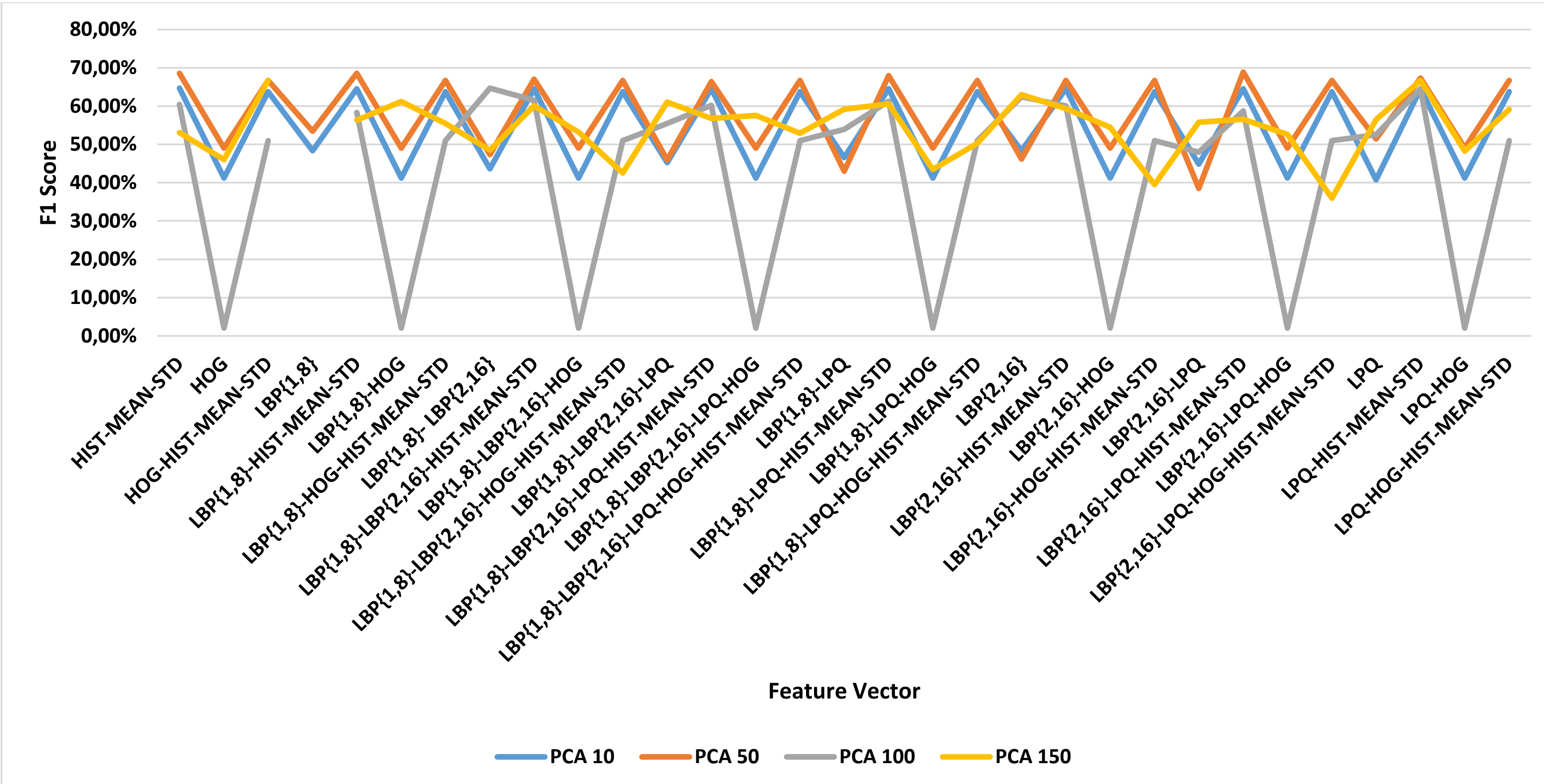


Figure 27 F1 Score for GMHI with Naive Bayes Classifier

5.3.1 Comparison of the Implemented Classifiers

Having tested several configuration set of parameters and comparing the classifier's performance, we noticed that feature combinations results do not outperform results of individual descriptors. Thus in the following comparison tables (Table 3 and Table 4) feature combination results are not included. The tables only include F1 score results of LBP{1,8}, LBP{2,16}, HOG, LPQ, HIST-

MEAN-STD, and feature fusion of all of the features together. For the MHI approach the best performance is 81.93% and is achieved with the appearance-based descriptor HOG and with combination of HIST-MEAN-STD using SVM classifier for 100 selected features.

	PCA 10				PCA 50				PCA 100				PCA 150			
Features	**N. Bayes**	**kNN**	**R. Forest**	**SVM**	**N. Bayes**	**kNN**	**R. Forest**	**SVM**	**N. Bayes**	**kNN**	**R. Forest**	**SVM**	**N. Bayes**	**kNN**	**R. Forest**	**SVM**
LBP{1,8}	45.50%	35.50%	42.71%	-	42.06%	29.94%	51.74%	-	-	-	-	-	-	-	-	-
LBP{2,16}	49.74%	29.09%	47.24%	-	44.10%	37.13%	48.68%	-	**64.34%**	27.21%	42.11%	-	45.03%	42.39%	44.44%	-
HOG	32.50%	38.51%	42.62%	79.80%	42.55%	38.37%	57.73%	25.45%	-	1.98%	57.58%	***81.93%***	53.59%	37.16%	54.74%	25.45%
LPQ	18.90%	**43.31%**	50.00%	60.20%	44.57%	49.02%	51.06%	**57.44%**	52.80%	31.08%	49.29%	59.30%	**62.45%**	42.78%	**56.46%**	**56.28%**
HIST-MEAN-STD	**64.08%**	37.57%	41.88%	**80.00%**	52.02%	48.94%	51.00%	25.45%	55.02%	**42.53%**	54.11%	***81.93%***	58.62%	**50.00%**	40.21%	25.45%
FEATURE FUSION	**64.08%**	38.37%	**54.73%**	54.55%	**64.15%**	**54.19%**	**61.22%**	38.10%	58.56%	36.84%	**61.14%**	36.59%	3,85%	23.84%	59.38%	48.36%

Table 3 F1 Score results of the implemented classifiers for MHI

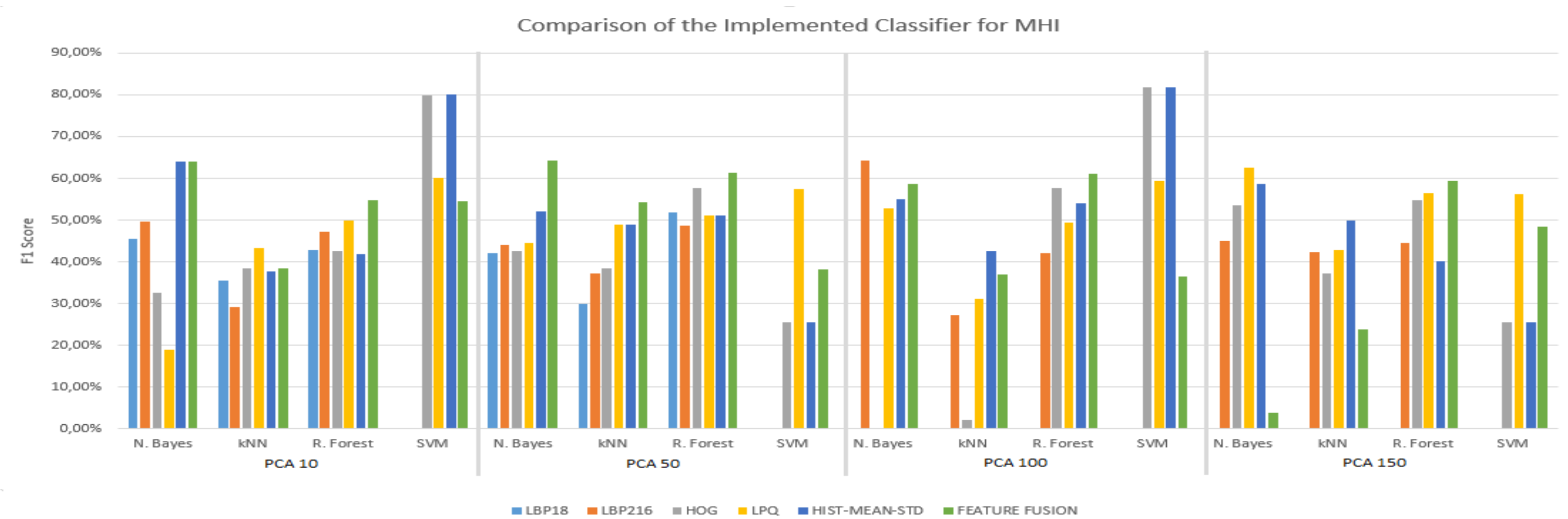


Figure 28 Comparison of the implemented Classifiers for MHI

As for GMHI approach the maximum F1 score is 81.93% achieved by the combination of statistical features HIST-MEAN-STD for 10 selected features.

	PCA 10				PCA 50				PCA 100				PCA 150			
Features	**N. Bayes**	**kNN**	**R. Forest**	**SVM**	**N. Bayes**	**kNN**	**R. Forest**	**SVM**	**N. Bayes**	**kNN**	**R. Forest**	**SVM**	**N. Bayes**	**kNN**	**R. Forest**	**SVM**
LBP{1.8}	48.31%	35.58%	52.53%	-	53.46%	47.56%	49.20%	-	-	-	-	-	-	-	-	-
LBP{2,16}	48.35%	42.50%	45.69%	-	46.15%	0.37%	42.64%	-	**62.41%**	40.52%	49.76%	-	**62.94%**	42.72%	56.4%	-
HOG	41.18%	**45.24%**	56.84%	62.72%	49.03%	38.75%	46.70%	25.45%	2.06%	2.02%	46.39%	**80.00%**	46.01%	46.86%	43.39%	25.45%
LPQ	40.76%	44.97%	45.03%	70.39%	51.46%	36.99%	48.45%	53.85%	52.54%	22.90%	48.70%	69.83%	56.48%	39.78%	48.04%	**54.55%**
HIST-MEAN-STD	**64.69%**	43.24%	**55.67%**	**81.93%**	**68.52%**	**57.73%**	52.68%	25.45%	60.40%	**57.73%**	**53.14%**	**80.00%**	53.06%	**61.03%**	**56.87%**	25.45%
FEATURE FUSION	63.77%	44.92%	55.00%	47.42%	66.67%	54.13%	**58.62%**	**61.47%**	50.98%	6.56%	51.65%	63.95%	52.88%	32.22%	54.82%	11.43%

Table 4 F1 Score results of the implemented classifiers for GMHI

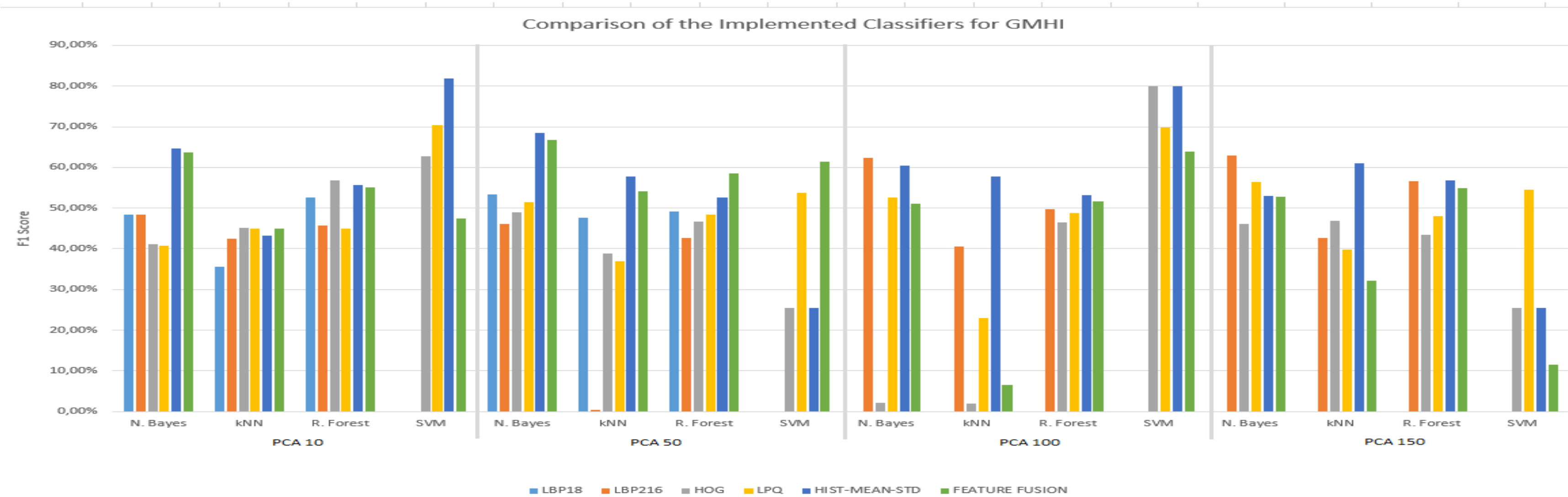


Figure 29 Comparison of the implemented Classifiers for GMHI

In summary, the MHI approach classifier Naïve Bayes results reached 64.34% F1 score, kNN 54.19%, and Random Forest 61.22%. For the GMHI approach Naïve Bayes classifier reached 64.69%, kNN 61.03% and Random Forest 58.62%. In both approaches the result of the above mentioned classifiers are much lower than the SVM classifier performance. However these classifiers remain promising for further exploration with different set of parameters.

LBP feature vector did not perform well in the overall. LBP{1,8} is a vector 1×59 thus for PCA 100 and 150 the algorithm cannot be executed. Also for SVM classifier in both MHI and GMHI approaches it presents null recognition. The maximum F1 score with LBP feature vector is achieved with Naïve Bayes classifier. Its performance may be improved by choosing different radius and neighborhood parameters. LPQ descriptor is another descriptor that its performance is not satisfactory as its best performance is 70.39% for the GMHI approach with SVM classifier. HIST-MEAN-STD and HOG performed equally well achieving the maximum overall algorithm performance. Figure 30 illustrates the best results for MHI, all implemented classifiers for PCA 100 and Figure 31 illustrates the best results for GMHI, all implemented classifiers for PCA 10.

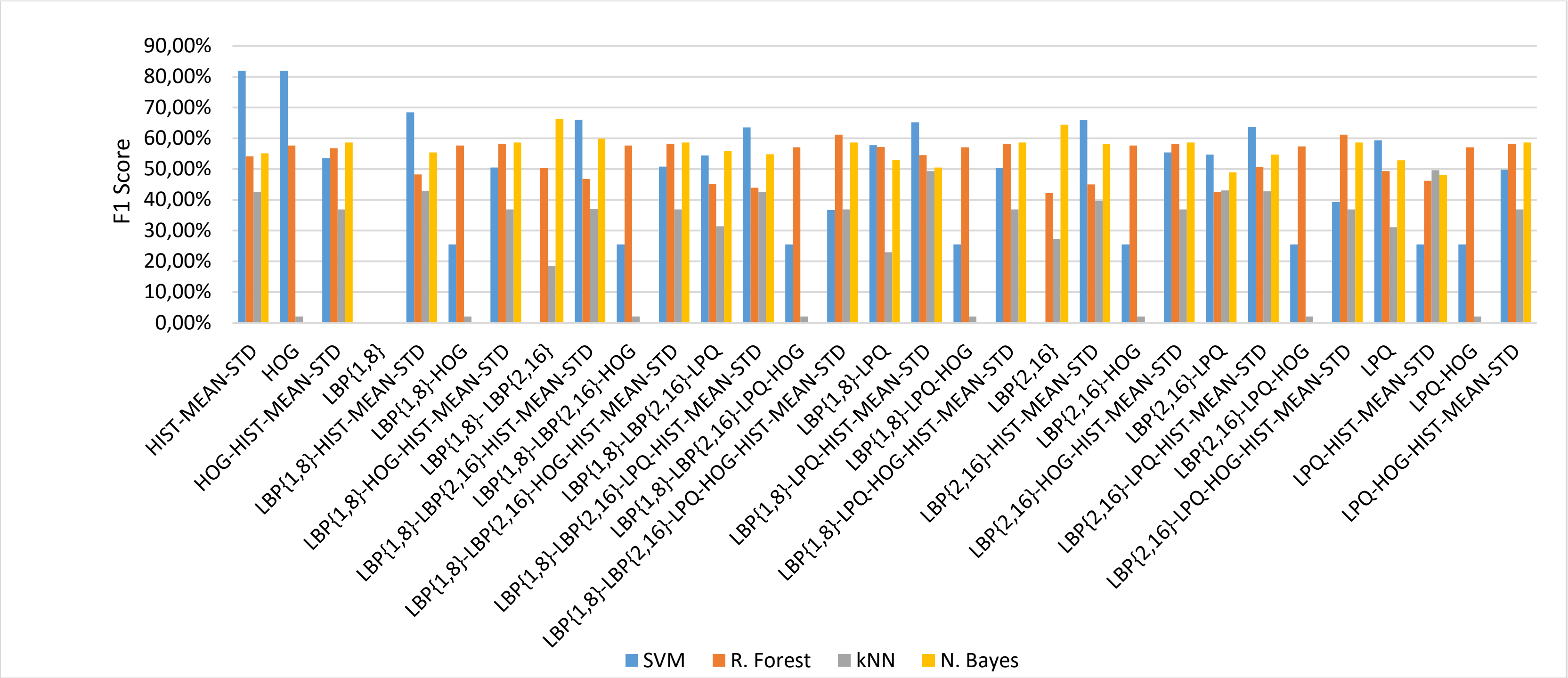


Figure 30 Performance of the Implemented Classifiers for MHI for all combinations of the feature vectors

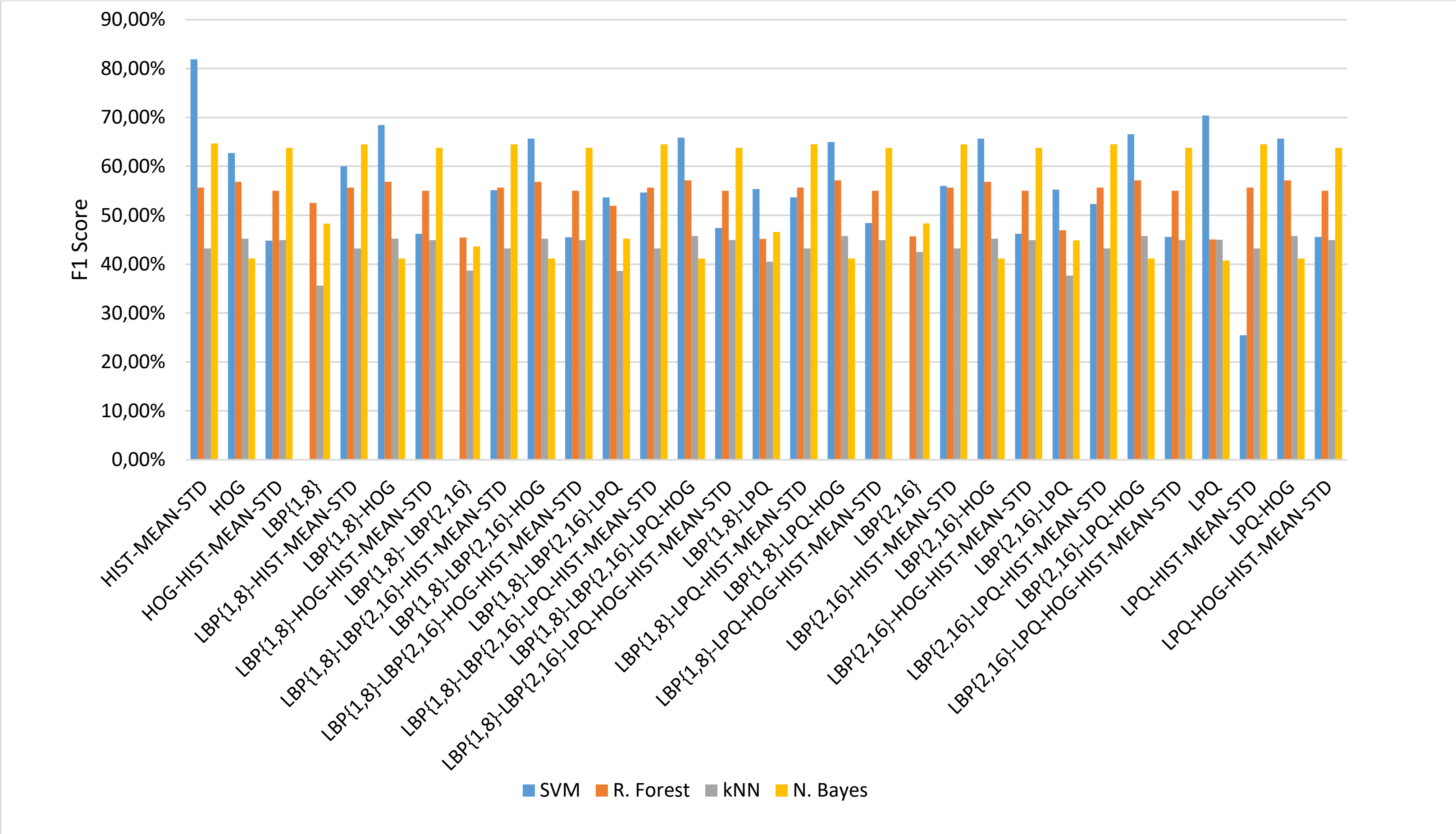
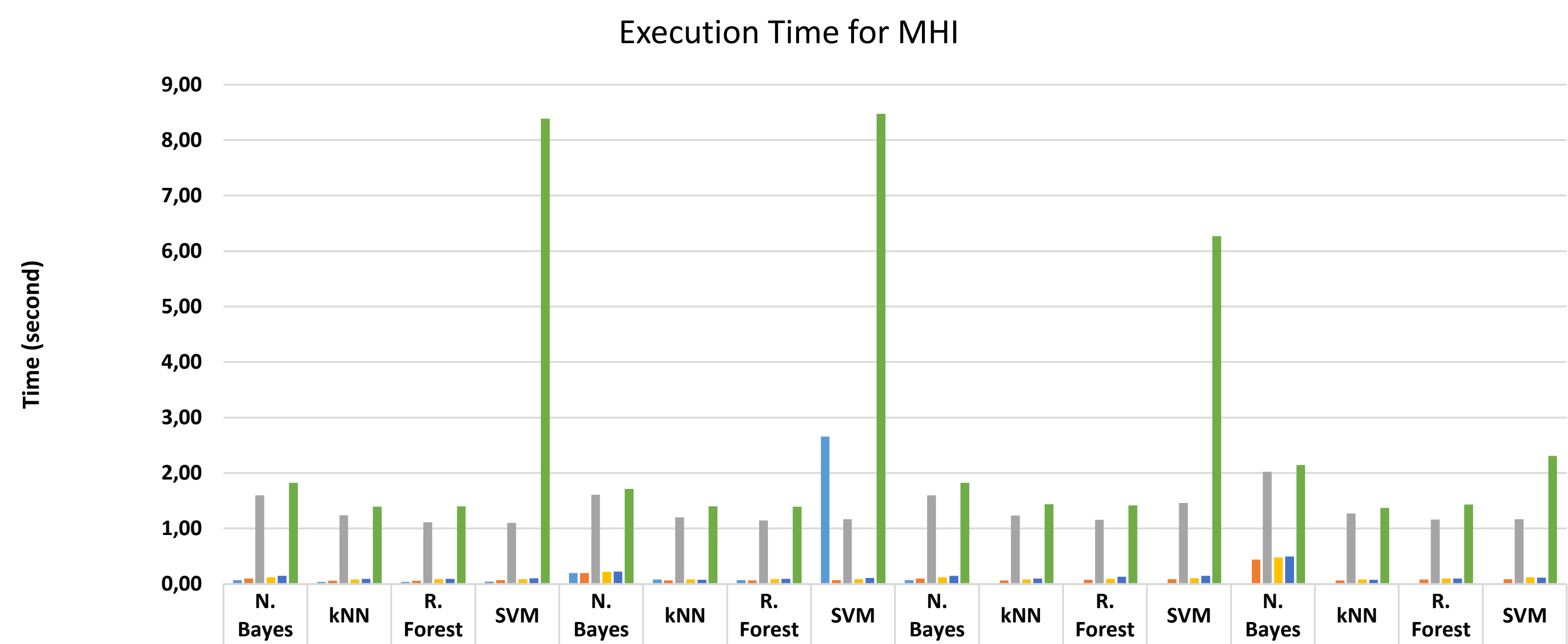


Figure 31 Performance of the Implemented Classifiers for GMHI for all combinations of the feature vectors

5.3.2 Execution Time

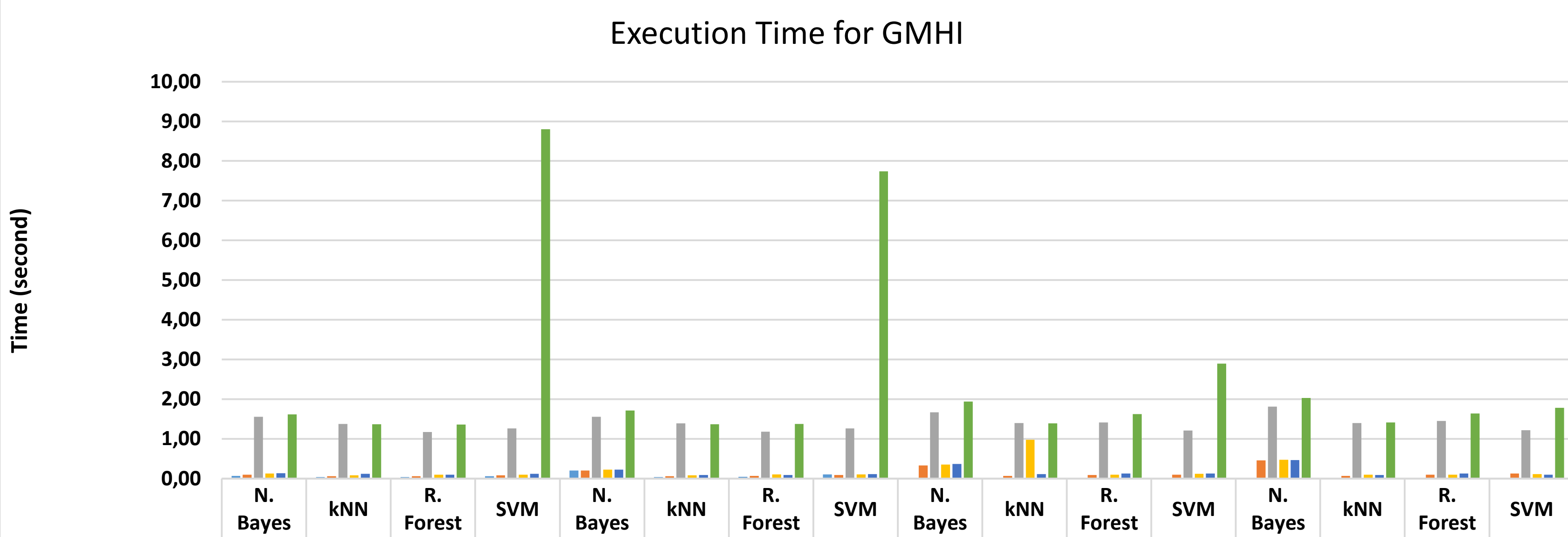
The execution time was also examined in the present work. This issue plays a significant role for the algorithm development. It would be inefficient to develop a slow algorithm with satisfactory results or a fast algorithm with poor results. Our aim is to select the optimal configuration set of parameters that will provide high accuracy with low execution time. The execution time depends on the characteristics of the system's setup. The experiments were tested on a system with on a 64-bit Operating System consisting of a processor Intel core i5-2500 CPU on 3.30GHz, and 10GB RAM. Figure 32 and Figure 33 include the execution time of the developed classifiers for MHI and GMHI approach respectively. We notice that GMHI approach is slightly faster than the MHI approach.



	N. Bayes	kNN	R. Forest	SVM	N. Bayes	kNN	R. Forest	SVM	N. Bayes	kNN	R. Forest	SVM	N. Bayes	kNN	R. Forest	SVM
LBP18	0,07	0,03	0,04	0,04	0,19	0,08	0,07	2,66	0,07	0,00		0,00				0,00
LBP216	0,09	0,06	0,06	0,07	0,20	0,06	0,07	0,07	0,09	0,06	0,07	0,09	0,44	0,06	0,08	0,09
HOG	1,60	1,24	1,11	1,10	1,61	1,20	1,15	1,16	1,60	1,23	1,15	1,46	2,02	1,27	1,16	1,16
LPQ	0,12	0,08	0,08	0,09	0,22	0,08	0,09	0,09	0,12	0,08	0,09	0,10	0,48	0,08	0,10	0,12
HIST-MEAN-STD	0,15	0,09	0,09	0,11	0,22	0,07	0,09	0,11	0,15	0,10	0,13	0,15	0,49	0,08	0,10	0,11
FEATURE FUSION	1,82	1,39	1,40	8,38	1,71	1,40	1,39	8,47	1,82	1,44	1,42	6,27	2,14	1,37	1,43	2,31

Figure 32 Execution Time for MHI

The maximum result 81.93% for MHI approach, HOG feature vector with SVM classifier and 100 PCA needs 1.46 seconds and HIST-MEAN-STD for the same parameters needs 0.15 seconds.The execution time variation of HOG feature vector and HIST-MEAN-STD feature vector is due to the value of the feature vector. HOG feature vector is 1×6084 and HIST-MEAN-STD is 255×1 so the execution time for HOG is greater than the HIST-MEAN-STD.



	N. Bayes	kNN	R. Forest	SVM	N. Bayes	kNN	R. Forest	SVM	N. Bayes	kNN	R. Forest	SVM	N. Bayes	kNN	R. Forest	SVM
LBP18	0,07	0,04	0,04	0,06	0,20	0,04	0,05	0,11		0,00				0,00		
LBP216	0,10	0,06	0,06	0,08	0,20	0,06	0,07	0,09	0,33	0,06	0,09	0,10	0,46	0,06	0,10	0,13
HOG	1,55	1,37	1,17	1,27	1,56	1,39	1,18	1,26	1,67	1,39	1,41	1,21	1,82	1,40	1,45	1,22
LPQ	0,12	0,08	0,09	0,10	0,22	0,08	0,10	0,10	0,35	0,98	0,10	0,12	0,47	0,09	0,10	0,11
HIST-MEAN-STD	0,13	0,12	0,09	0,12	0,22	0,09	0,09	0,11	0,37	0,11	0,13	0,12	0,46	0,09	0,12	0,10
FEATURE FUSION	1,62	1,36	1,36	8,80	1,71	1,37	1,37	7,74	1,94	1,39	1,62	2,90	2,03	1,42	1,64	1,78

Figure 33 Execution Time for GMHI

The maximum result 81.93% for GMHI approach, HIST-MEAN-STD feature vector with SVM classifier and 10 PCA needs 0.12 seconds.

6. Conclusion

Facial expressions convey significant information for the affective state of an individual. In summary the aim of this thesis was to develop an automated application for facial image analysis. The primary purpose is to detect visual signs of depression. The proposed work is based on the hypothesis that depressed individuals exhibit restricted facial motion than non-depressed The AVEC dataset was used for the system development. Alighted face images are extracted through the OpenFace toolkit. Face is the region of interest where two motion representation methods are implemented Motion History Image and Gabor Motion History Image where moving parts of a video sequence can be engraved in a single image. Low level appearance based descriptors were developed (Local Binary Pattern, Local Phase Quantization and Histogram of Oriented Gradients) as well as same statistical features (Mean standard deviation and combined histogram). Several classification algorithms were utilized in order to examine the sensitivity and specificity of the proposed algorithm. Naïve Bayes, k-Nearest Neighbors, Random Forest and Support Vector Machine classification algorithms were used. Support Vector Machine classifier achieved the best performance for both MHI and GMHI motion representation methods. SVM classifier is widely used in categorical assessment of depression. It seems to perform satisfying results for binary problems with high dimensionality [11].

In the overall, and with respect to the research questions that were set in section 1.1, the work conducted in terms of this thesis, was able to provide some answers. More specifically for the MHI approach the best performance 81.93% achieved by appearance-based descriptor HOG and with combination of HIST-MEAN-STD using SVM classifier for 100 selected features. The execution time for those results are 1.46 seconds and 0.15 seconds respectively. Consequently, the HIST-MEAN-STD feature vector with 100 PCA for MHI maximize the overall system performance 81.93% with the minimum execution time. As for GMHI approach, the maximum F1 score is 81.93% achieved by the combination of statistical features HIST-MEAN-STD for 10 selected features and the execution time is 0.12 seconds.

6.1 Future work

The proposed framework introduced a novel variant of MHI, the Gabor Motion History Image. However, there is still room for further exploration. A multimodal approach of the implemented framework would be desirable, by incorporating audio-based features. Another expansion could be to do more tests with additional classifiers and other low level features, which could improve the overall performance of the algorithm. Furthermore, the proposed framework could be implemented in different dataset.

Bibliography

[1] European Statistical System, "Mental health and related issues statistics," EUROSTAT, 2017.

[2] C. Mathers and World Health Organization, The Global Burden of Disease: 2004 Update, World Health Organization, 2008.

[3] World Health Organization, World health statistics 2012, World Health Organization, 2012.

[4] Directorate-General and Healthe & Consumer Protection , "Green Paper Improving the mental health of the population: Towards a strategy on mental health for the European Union," European Communities, Brussels, 2005.

[5] S. B. Guze and E. Robins, "Suicide and primary affective disorders," *The British Journal of Psychiatry,* vol. 117, no. 539, pp. 437-438, 1970.

[6] World Health Organization, "Depression," February 2017. [Online]. Available: http://www.who.int/mediacentre/factsheets/fs369/en/. [Accessed August 2017].

[7] American Psychiatric Association, Diagnostic and Statistical Manual of Mental Disorders 4th Edition, Washington DC, USA: APA, 1994.

[8] H. Dibeklioğlu, Z. Hammal, Y. Yang and J. Cohn, "Multimodal Detection of Depression in Clinical Interviews," in *Proc ACM Int Conf Multimodal Interact*, 2015.

[9] A. T. Beck, C. H. Ward, M. Mendelson, J. Mock and J. Erbaugh, "An Inventory for Measuring Depression," *Archives of general Psychiatry,* vol. 4, no. 6, pp. 561-571, 1961.

[10] M. Hamilton, "A Rating Scale for Depression," *Journal of Neurology Neurosurgery and Psychiatry,* vol. 23, no. 1, pp. 56-62, February 1960.

[11] A. Pampouchidou, P. Simos, K. Marias, F. Meriaudeau, F. Yang, M. Pediaditis and M. Tsiknakis, "Automatic Assessment of Depression Based on Visual Cues: A Systematic Review," *IEEE Transactions on Affective Computing,* no. 99, 28 September 2017.

[12] M. Valstar, S. Schnieder, B. Schuller, B. Jiang, R. Cowie, K. Smith, S. Bilakhia and M. Pantic, "AVEC 2013: the continuous audio/visual emotion and depression recognition challenge," in *Proceedings of the 3rd ACM international workshop on Audio/visual emotion challenge*, Barcelona, Spain, 2013.

[13] M. Valstar, B. Schuller, K. Smith, T. Almaev, F. Eyben, J. Krajewski, R. Cowie and M. Pantic , "AVEC 2014: 3D Dimensional Affect and Depression Recognition Challenge," in *Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge*, Orlando, Florida, USA, 2014.

[14] M. Valstar, J. Gratch, B. Schuller, F. Ringeval, D. Lalanne, M. Torres, S. Scherer, G. Stratou, R. Cowie and M. Pantic, "AVEC 2016: Depression, Mood, and Emotion Recognition Workshop and

Challenge.," *AVEC '16 Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge,* pp. 3-10, 16 October 2016.

[15] A. Pampouchidou, O. Simantiraki, C. -. M. Vazakopoulou, C. Chatzaki, M. Pediaditis, A. Maridaki, K. Marias, P. Simos, F. Yang, F. Meriaudeau and M. Tsiknakis, "Facial Geometry and Speech Analysis for depression Detection," in *39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'17)*, Seogwipo, South Korea, 2017.

[16] A. PampouchidouEmail , M. Pediaditis, A. Maridaki, M. Awais, C.-M. Vazakopoulou, S. Sfakianakis, M. Tsiknakis, P. Simos, K. Marias, F. Yang and F. Meriaudeau, "Quantitative comparison of motion history image variants for video-based depression assessment," *EURASIP Journal on Image and Video Processing,* 2017.

[17] American Psychiatric Association, Diagnostic and Statistical Manual of Mental Disorders, DSM-5, Washington, DC, 2013.

[18] E. W. Copeland, A. Angold, J. E. Costello and H. Egger, "Prevalence, Comorbidity, and Correlates of DSM-5 Proposed Disruptive Mood Dysregulation Disorder," *The American Journal of Psychiatry,* vol. 170, no. 2, pp. 173-179, 01 February 2013.

[19] A. Rapkin, "A review of treatment of premenstrual syndrome & premenstrual dysphoric disorder," *Psychoneuroendocrinology,* vol. 28, no. 3, pp. 39-53, August 2003.

[20] J. B. W. Williams, "A structured interview guide for the Hamilton Depression Rating Scale.," *Archives of general psychiatry,* 1 August 1988.

[21] A. Shahid, K. Wilkinson, S. Marcu and C. M. Shapiro, STOP, THAT and One Hundred Other Sleep Scales, New York: Springer, 2012.

[22] A. P. Association, "DSM–5: Frequently Asked Questions," American Psychiatric Association, [Online]. Available: https://www.psychiatry.org/psychiatrists/practice/dsm/feedback-and-questions/frequently-asked-questions. [Accessed 4 August 2017].

[23] D. C. Răzvan , "Non-verbal Communication - Indispensable Complement of Oral and Written Communication," *Procedia - Social and Behavioral Sciences,* vol. 137, pp. 105-111, 09 July 2014.

[24] Y. Tian, L. Brown, A. Hampapur, S. Pankanti, A. Senior and R. Bolle, "Real world real-time automatic recognition of facial expression," in *In Proceedings of IEEE workshop*, Australia, 2003.

[25] P. Ekman and E. Rosenberg, What the face reveals, New York: Oxford University Press, 2005.

[26] J. F. McGowan and L. D. Schmidt, Counselling: Readings in Theory and Practice, New York: Holt,Rinehart & Winston of Canada Ltd, 1962.

[27] P. H. Waxer, "Nonverbal cues for depression," *Journal of Abnormal Psychology,* vol. 83, no. 3, pp. 319-322, 1974.

[28] W. Gaebel and W. Wölwer, "Facial expression and emotional face recognition in schizophrenia and depression," *European Archives of Psychiatry and Clinical Neuroscience,* vol. 242, no. 1, pp. 46-52, September 1992.

[29] P. H. Waxer, "Therapist training in nonverbal communication i: Nonverbal," *Journal of Clinical Psychology,* vol. 30, no. 2, pp. 215-218, 1974.

[30] P. Ekman and W. V. Friesen, "Nonverbal behavior and psychopathology," *The psychology of depression Contemporary theory and research,* pp. 203-232, 1974.

[31] P. Ekman, W. V. Friesen, M. O'Sullivan, A. Chan, I. Diacoyanni - Tarlatzis, K. Heider, R. Krause, W. A. LeCompte, T. Pitcaim, P. E. Ricci-Bitti, K. Scherer, M. Tomita and A. Tzavaras, "Universals and cultural differences in the judgments of facial expressions of emotion," *Journal of Personality and Social Psychology,* vol. 53, no. 4, pp. 712-717, 1987.

[32] P. Ekman, D. Matsumoto and W. V. Friesen, "Facial Expression in Affective Disorders," *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS),* pp. 429-439, 1997.

[33] P. Ekman and R. Davidson, The nature of emotion: Fundamental questions, New York: Oxford University Press, 1994.

[34] S. Alghowinem, R. Goecke, M. Wagner, G. Parker and M. Breakspear, "Eye Movement Analysis for Depression Detection," in *International Conference on Image Processing*, 2013.

[35] S. Scherer, G. Stratou and L.-P. Morency, "Audiovisual Behavior Descriptors for Depression Assessment," *Image and Vision Computing,* vol. 32, no. 10, pp. 648-658, 2014.

[36] R. Picard, Affective Computing, Massachusett: MIT Press, 1997.

[37] S. Phimoltares, C. Lursinsap and K. Chamnongthai, "Face detection and facial feature localization without considering the appearance of image context," *Image and Vision Computing,* vol. 25, pp. 741-753, 2007.

[38] C. A. Corneanu, M. O. Simon and J. F. Cohn, "Survey on RGB, 3D, Thermal, and Multimodal Approaches for Facial Expression Recognition: History, Trends, and Affect-Related Applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 38, no. 8, pp. 1548 - 1568, 1 August 2016.

[39] Ying-Li Tian, Takeo Kanade and Jeffrey F. Cohn, "Facial Expression Analysis," in *Handbook of Face Recognition*, New York, NY, Springer, 2005, pp. 247-275.

[40] A. Pampouchidou, O. Simantiraki, A. Fazlollahi, M. Pediaditis, D. Manousos, A. Roniotis, G. Giannakakis, F. Meriaudeau, P. Simos, K. Marias, F. Yang and M. Tsiknakis, "Depression Assessment by Fusing High and Low Level Features from Audio, Video, and Text," *Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge,* pp. 27-34, 16 October 2016.

[41] James R. Williamson, Thomas F. Quatieri, Brian S. Helfer, Rachelle Horwitz, Bea Yu and Daryush D. Mehta , "Vocal biomarkers of depression based on motor incoordination," in *AVEC '13 Proceedings of the 3rd ACM international workshop on Audio/visual emotion challenge*, Barcelona, Spain, 2013.

[42] Lu-Shih Alex Low, Namunu C. Maddage and Margaret Lech, "Influence of acoustic low-level descriptors in the detection of clinical depression in adolescents," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, Dallas, TX, USA, 2010.

[43] Namunu C Maddage, Rajinda Senaratne, Lu-Shih Alex Low, Margaret Lech and Nicholas Allen, "Video-based detection of the clinical depression in adolescents," in *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*, Minneapolis, MN, USA, 2009.

[44] B. Bediou, P. Krolak-Salmon, M. Saoud, M.-A. Hanaff, M. Burt, J. Dalery and T. D'Amato, "Facial Expression and Sex Recognition in Schizophrenia and Depression," *The Canadian Journal of Phychiatry,* vol. 50, no. 9, pp. 525-533, 1 August 2005.

[45] Yu Zhu, Yuanyuan Shang, Zhuhong Shao and Guodong Guo , "Automated Depression Diagnosis based on Deep Networks to Encode Facial Appearance and Dynamics," *IEEE Transactions on Affective Computing,* 10 January 2017.

[46] Sharifa Alghowinem, Roland Goecke, Michael Wagner, Julien Epps, Matthew Hyett, Gordon Parker and Michael Breakspear, "Multimodal depression detection: fusion analysis of paralinguistic, head pose and eye gaze behaviors," *IEEE Transactions on Affective Computing,* 21 January 2016.

[47] Nicholas Cummins, Jyoti Joshi, Abhinav Dhall, Vidhyasaharan Sethu, Roland Goecke and Julien Epps, "Diagnosis of Depression by Behavioural Signals: A Multimodal Approach," in *AVEC '13 Proceedings of the 3rd ACM international workshop on Audio/visual emotion challenge*, Barcelona, Spain, 2013.

[48] Maxim Sidorov and Wolfgang Minker, "Emotion Recognition and Depression Diagnosis by Acoustic and Visual Features: A Multimodal Approach," in *AVEC '14 Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge*, Orlando, Florida, USA, 2014.

[49] T. Kanade, Picture Processing System by Computer Complex and recognition of human faces, Kyoto, 1973.

[50] P. Ekman, T. S. Huang, T. J. Sejnowski and J. C. Hager, Final Report To NSF of the Planning Workshop on Facial Expression Understanding, Washington DC: National Science Foundation, 1992.

[51] J. M. Girard and J. F. Cohn, "Automated audiovisual depression analysis," *Current Opinion in Psychology,* vol. 4, pp. 75-79, August 2015.

[52] P. Ekman and W. Friesen, Facial Action Coding System, Consulting Psychologists Press, 1978.

[53] M. A. R. Ahad, J. K. Tan, H. Kim and S. Ishikawa, "Motion history image: its variants and applications," *Machine Vision and Applications,* pp. 255-281, March 2012.

[54] A. F. Bobick and J. W. Davis, "Real-time Recognition of Activity Using Temporal Templates," in *3rd IEEE Workshop on Applications of Computer Vision, 1996*, Sarasota, FL, USA, USA, 1996.

[55] M. A. R. Ahad, Motion history images for action recognition and understanding, London: Springer, 2013.

[56] M. Valstar, M. Pantic and I. Patras, "Motion History for Facial Action Detection in Video," in *IEEE International Conference on Systems, Man and Cybemetics*, The Hague, Netherlands, 2004.

[57] H. Meng, D. Huang, H. Wang, H. Yang, M. AI-Shuraifi and Y. Wang, "Depression recognition based on dynamic facial and vocal expression features using partial least square regression," *AVEC '13 Proceedings of the 3rd ACM international workshop on Audio/visual emotion challenge,* pp. 21-30, 21 October 2013.

[58] A. Jan, H. Meng, Y. F. A. Gaus, F. Zhang and S. Turabzadeh, "Automatic Depression Scale Prediction using Facial Expression Dynamics and Regression," *AVEC '14 Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge,* pp. 73-80, 07 November 2014 .

[59] B. Manjunath, C. Shekhar and R. Chellappa, "A new approach to image feature detection with applications," *Pattern Recognition,* vol. 29, no. 4, pp. 627-640, April 1996.

[60] M. Lyons, S. Akamatsu and M. Kamachi, "Coding facial expressions with Gabor wavelets," *Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition,* April 1998.

[61] B. Fasel and J. Luettin, "Automatic facial expression analysis: a survey," *Pattern Recognition,* vol. 36, no. 1, pp. 259-275, January 2003.

[62] J. M. Girard, . J. F. Cohn, M. H. Mahoor, S. M. Mavadati, Z. Hammal and D. P. Rosenwald, "Nonverbal Social Withdrawal in Depression: Evidence from manual and automatic analysis," *Image and Vision Computing,* vol. 32, no. 10, pp. 641-647, 2013.

[63] J. M. Girard, Jeffrey F. Cohn, Mohammad H. Mahoor, Seyedmohammad Mavadati and Dean P. Rosenwald, "Social Risk and Depression: Evidence from Manual and Automatic Facial Expression Analysis," *in 10th International Conference on Automatic Face and Gesture Recognition,* 2013.

[64] Namunu C Maddage, Rajinda Senaratne, Lu-Shih Alex , Margaret Lech and Nicholas Allen, "Video-Based Detection of the Clinical Depression in Adolescents," *31st Annual International Conference of the IEEE EMBS,* pp. 3723-3726, September 2009.

[65] A. Cruz, B. Bhanu and N. S. Thakoor, "Facial Emotion Recognition with Anisotropic," in *IEEE International Conference on Image Processing*, Melbourne, Victoria, Australia, 2013.

[66] C. Grigorescu, N. Petkov and M. A. Westenberg, "Contour Detection Based on Nonclassical Receptive Field Inhibition," *IEEE Transactions on Image Processing,* vol. 12, no. 7, pp. 729-739, 15 july 2003.

[67] N. Dalal, B. Triggs and C. Schmid, "Human Detection Using Oriented Histograms of Flow and Appearance," in *9th European Conference on Computer Vision*, Graz, Austria, 2006.

[68] Haoyu Ren and Ze-Nian Li, "Object detection using edge histogram of oriented gradient," in *2014 IEEE International Conference on Image Processing (ICIP)*, Paris, France, 2014.

[69] Le Yang, Dongmei Jiang, Lang He, Ercheng Pei, Meshia Cédric Oveneke and Hichem Sahli, "Decision Tree Based Depression Classification from Audio Video and Language Information," in *AVEC '16 Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge*, Amsterdam, The Netherlands, 2016.

[70] J. H. Ville Ojansivu, "Blur Insensitive Texture Classification Using Local Phase Quantization," in *ICISP 2008: Image and Signal Processing*, Berlin, Heidelberg, 2008.

[71] Bihan Jiang, Michel F. Valstar and Maja Pantic, "Action unit detection using sparse appearance descriptors in space-time video volumes," in *Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition*, Santa Barbara, CA, USA, 2011.

[72] M. Kachele, M. Glodek, D. Zharkov, S. Meudt and F. Schwenker, "Fusion of Audio-visual Features using Hierarchical Classifier Systems for the Recognition of Affective States and the State of Depression," in *3rd International Conference on Pattern Recognition Applications and Methods (ICPRAM)*, Loire Valley, France, 2014.

[73] Y. Zhu, Yuanyuan Shang and Zhuhong Shao, "Automated Depression Diagnosis based on Deep Networks to Encode Facial Appearance and Dynamics," *IEEE Transactions on Affective Computing,* 10 January 2017.

[74] T. Ojala, M. Pietikäinen and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognition,* vol. 29, no. 1, pp. 51-59, January 1966.

[75] Timo Ahonen, Abdenour Hadid and Matti Pietikäinen, "Face Recognition with Local Binary Patterns," in *Computer Vision - ECCV 2004*, Verlag Berlin Heidelberg, Springer, 2004, pp. 469-481.

[76] C. Shan, S. Gong and P. W. McOwan, "Robust facial expression recognition using local binary patterns," in *IEEE International Conference on Image Processing 2005*, Genova, Italy, 2005.

[77] R. Gupta, N. Malandrakis, B. Xiao, T. Guha, V. M. Segbroeck, M. Black, A. Potamianos and S. Narayanan, "Multimodal Prediction of Affective Dimensions and Depression in Human-Computer Interactions," in *4th International Workshop on Audio/Visual Emotion Challenge*, Orlando, USA, 2014.

[78] Bihan Jiang, Michel F. Valstar and Maja Pantic, "Action unit detection using sparse appearance descriptors in space-time video volumes," in *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, Santa Barbara, CA, USA, USA, 2011.

[79] J. Joshi, "An Automated Framework for Depression Analysis," in *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, Geneva Switzerland, 2013.

[80] J. Gratch, R. Artstein, G. Lucas, G. Stratou, S. Scherer, A. Nazarian, R. Wood, J. Boberg, D. DeVault, S. Marsella, D. Traum, S. Rizzo and L.-P. Morency, "The Distress Analysis Interview Corpus of human and computer interviews," in *Proceedings of the Ninth International Conference on Language Resources and Evaluation* , Reykjavik, Iceland, 2014.

[81] B. Löwe, K. Kroenke, W. Herzog and K. Gräfe, "Measuring depression outcome with a brief self-report instrument: sensitivity to change of the Patient Health Questionnaire (PHQ-9)," *Journal of Affective Disorders,* vol. 81, no. 1, pp. 61-66, july 2001.

[82] Mohammed Senoussaoui, Milton Sarria-Paja, João F. Santos and Tiago H. Falk, "Model Fusion for Multimodal Depression Classification and Level Detection," in *AVEC '14 Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge*, Orlando, Florida, USA, 2014.

[83] Tadas Baltrušaitis, Peter Robinson and Louis-Philippe Morency, "OpenFace: An open source facial behavior analysis toolkit," in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, Lake Placid, NY, USA, 2016.

[84] Tadas Baltrusaitis, Peter Robinson and Louis-Philippe Morency, "Constrained Local Neural Fields for robust facial landmark detection in the wild," in *IEEE International Conference on Computer Vision Workshops*, 2013.

[85] David Cristinacce and Timothy F. Cootes, "Feature Detection and Tracking with Constrained Local Models," in *Proceedings of the British Machine Vision Conference*, Edinburgh, UK, 2006.

[86] Aaron F. Bobick and James W. Davis, "The Recognition of Human Movement Using Temporal Templates," *IEEE Transaction on Pattern Analysis and Machine Intelligence,* vol. 23, no. 3, pp. 257-267, March 2001.

[87] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, San Diego, CA, USA, 2005.

[88] Oren Barkan, Jonathan Weill, Lior Wolf and Hagai Aronowitz, "Fast High Dimensional Vector Multiplication Face Recognition," in *2013 IEEE International Conference on Computer Vision*, Sydney, NSW, Australia, 2013.

[89] Di Huang, Caifeng Shan, Mohsen Ardabilian, Yunhong Wang and Liming Chen, "Local Binary Patterns and Its Application to Facial Image Analysis: A Survey," *IEEE Transactions on Systems, Man, and Cybernetics, Part C,* pp. 765-781, 28 March 2011.

[90] I. Guyon, S. Gunn, M. Nikravesh and L. Zadeh, Feature Extraction: Foundation and Applications, New York: Springer, 2006.

[91] S.B. Kotsiantis, "Supervised Machine Learning: A Review of Classification Techniques," in *Emerging Artificial Intelligence Applications in Computer Engineering: eal Word AI Systems with Applications in EHealth, HCI, Information Retrieval and Pervasive Technologies*, IOS Press, 2007, pp. 249-268.

[92] M.Pal, "Random forest classifier for remote sensing classification," *International Journal of Remote Sensing,* pp. 217-222, 22 January 2007.

[93] Taiwo Oladipupo Ayodele, "Types of Machine Learning Algorithms," in *New Advances in Machine Learning*, INTECH. Science, Technology and Medicine open access publisher, 2010.

[94] Ron Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *IJCAI'95 Proceedings of the 14th international joint conference on Artificial intelligence*, Montreal, Quebec, Canada, 1995.

[95] Max Bramer, "Measuring the Performance of a Classifier," in *Principles of Data Mining*, London, Springer, 2013, pp. 175-178.