

Contents

List of Figures	xv
List of Tables	xix
Abbreviations	xxi
Symbols	xxv
Preface	xxvii
1 Introduction	1
1.0.1 The Role of Electromyography in Hand Gesture Recognition for Intuitive HCI	2
1.0.2 surface Electromyography(sEMG)	3
1.0.2.1 EMG Signals: Mathematical models based on contractions	4
1.0.2.2 Limitations of Contraction Models	4
1.0.3 Motivation	5
1.0.4 Problem Statement	7
1.1 Thesis Objectives	8
1.1.1 Mapping Electromyographic (EMG) Signals to Hand Gestures (Static and Dynamic)	8
1.1.1.1 Collecting sEMG Signals and Signal Preprocessing	9
1.1.1.2 Feature Engineering: Feature ensemble and meta-heuristic feature selection	9
1.1.1.3 Hand Gesture Recognition: Classification Using Machine Learning Methods	10
1.1.2 Dynamic Hand Gesture Recognition: sEMG-Based approach for Handwritten Character Recognition (HCR)	11
1.1.3 Multi-Modal Handwritten Character Recognition	11
1.1.4 Efficient hand grasp recognition using a minimal number of sEMG signals	11
1.1.5 Enhancing Interpretability with Explainable AI (XAI)	12

1.2	Contribution of thesis	13
2	Related Work and Preliminaries	19
2.1	EMG-based Sign Language Gesture Recognition	20
2.2	EMG-based handwritten character gesture recognition	23
2.2.1	A Multi-Modal Sensors Feature Fusion Approach for Handwritten Characters Recognition	27
2.2.1.1	Handwriting dynamics as a diagnostic tool for neurological diseases	31
2.3	EMG-Based Prosthetic Hand Grasp Recognition	34
2.3.1	Other insightful observations	36
2.3.1.1	Data segmentation	36
2.3.1.2	Response Time of Hand Gesture Recognition Models	37
2.3.1.3	Data acquisition device	37
2.3.2	Preliminaries	38
2.3.2.1	Myo armband: wearable sensor used for data collection	38
2.3.3	Datasets	39
2.3.4	Dataset-I:Ninapro reference dataset(Publicly available)	39
2.3.5	Dataset-II: Hand grasp dataset(Publicly available)	40
2.3.6	Dataset-III: American Sign Language digit dataset, ASL-10 (Collected Dataset)	41
2.3.7	Dataset-IV: ASL manual alphabet dataset, ASL-24 (Collected Dataset)	41
2.3.8	Dataset-V: sEMG based Handwritten character recognition dataset, S-HCR (Collected dataset)	44
2.3.9	Dataset-VI: Multi-Modal Handwritten Character Recognition dataset, MM-HCR (collected dataset)	45
2.3.10	Dataset-VII: DARWIN, used for Alzheimer prediction (Publicly available)	45
2.3.11	Dataset-VIII: UCI datasets: For validating proposed meta-heuristic feature selection approach	47
2.3.12	Evaluation Metrics	48
2.3.12.1	Accuracy	48
2.3.12.2	Matthews Correlation Coefficient (MCC)	49
2.3.12.3	Kappa Score	49
2.3.12.4	Receiver Operating Characteristic (ROC) Curve	50
2.3.12.5	Response Time	50
2.3.12.6	Explainable AI (XAI): Leveraging SHAP (SHapley Additive exPlanations) for Interpreting Machine Learning Models	51
2.4	Feature Selection Algorithms	52
2.4.1	ReliefF Algorithm	53
2.4.1.1	Equations	54
2.4.1.2	Handling Incomplete Data	54

2.4.1.3	Probabilistic Interpretation	54
2.4.2	Analysis of Variance (ANOVA) based feature selection	54
2.4.2.1	Equations and Variables	55
2.4.2.2	Variable Explanations	56
2.4.2.3	Interpretation	56
2.4.3	Chi-square for Feature Selection	57
2.4.3.1	Equations and Variables	58
2.4.3.2	Variable Explanations	58
2.4.3.3	Interpretation	59
2.4.4	Mutual Information for Feature Selection	59
2.4.4.1	Equations and Variables	60
2.4.4.2	Variable Explanations	60
2.4.4.3	Interpretation	61
2.5	Tools and Frameworks	61
2.5.1	Keras	61
2.5.2	Tensorflow	62
2.5.3	Raspberry Pi 3	62
3	Enhancing sEMG-Based Static Hand Gesture Recognition Using Machine Learning Approach	63
3.1	Introduction	64
3.1.1	Problem Definition and overview of the solution	67
3.1.1.1	Problem Statement	67
3.1.1.2	Overview of the Solution	67
3.1.2	Major contribution of the work	70
3.2	Materials and Methods	71
3.2.1	Subjects and experimental protocol	72
3.2.2	Data Acquisition and Sensor Placement	72
3.2.3	Datasets	74
3.2.3.1	Dataset Preparation	75
3.2.4	Data Preprocessing	75
3.2.4.1	Filters	76
3.2.5	Data Segmentation	76
3.2.6	Feature Extraction	77
3.2.7	Proposed Ensemble feature selection approach	77
3.2.7.1	Overall Complexity of the Ensemble Feature Selection Algorithm	79
3.3	Results and discussion	83
3.3.1	Performance Measure	84
3.3.2	Validation of proposed framework	88
3.3.2.1	Statistical Analysis	88
3.3.2.2	5*2 CV t-test	91
3.3.2.3	Model Scalability and Training Sample Sufficiency	92
3.3.2.4	Validation using public dataset: NinaPro [1]	94

3.3.3	Comparison to state-of-the-art methods	96
3.3.4	Insightful observations on selected features	98
3.3.4.1	Feature visualization	101
3.3.4.2	Others Insights on Feature using Explainable AI (SHAP [2])	101
3.3.5	Run Time Analysis	104
3.3.6	Threats to Validity	105
3.4	Summary	106
4	Leveraging cooperative theory based feature selection for hand grasp classification using the minimal number of sEMG signals	109
4.1	Introduction	109
4.1.1	Problem Statement and overview of the solution	110
4.1.1.1	Problem Statement	110
4.1.1.2	Methodology for the Proposed Solution	111
4.1.2	Major contribution of the work	111
4.2	Materials and methods	113
4.2.1	Dataset	113
4.2.2	Data Preprocessing	113
4.2.3	Feature Extraction	113
4.3	Feature selection	114
4.3.1	Preliminaries concepts and proposed method	115
4.3.1.1	Cooperative game	115
4.3.1.2	SHAP (SHapley Additive exPlanations)	118
4.3.2	Proposed feature Selection Algorithm	120
4.4	Results and discussions	124
4.4.1	Performance Measure	124
4.4.2	State-of-the-Art comparison	127
4.4.3	Model Scalability and Training Sample Sufficiency	129
4.4.4	Insights on Significant Feature Contributions	131
4.4.4.1	Friedrich coefficients	131
4.4.4.2	Fourier entropy	133
4.4.4.3	Matrix profile	133
4.4.4.4	IoT based Real-World Application Scenario	133
4.4.4.5	Threats to validity	134
4.5	Summary	135
5	Leveraging Deep feature learning for sEMG-based dynamic hand gesture recognition	139
5.1	Introduction	139
5.1.1	Problem Statement and Overview of the Solution	141
5.1.1.1	Problem Statement	141
5.1.1.2	Overview of the Solution	142
5.1.2	Major contribution of the work	144

5.2	Materials and methods	144
5.2.1	Sensor: Surface electromyography(sEMG)	144
5.2.1.1	Data collection device	145
5.2.2	Number of subjects and experimental protocol	145
5.2.3	Dataset preparation (S-HCR)	146
5.2.3.1	Filters	147
5.2.3.2	Data segmentation	147
5.2.4	Feature extraction	148
5.3	Preliminary Concepts and Proposed Architecture	150
5.3.1	Autoencoder	151
5.3.2	Stacked Autoencoder	152
5.3.3	Stacked Sparse Auto-encoder (SSAE)	153
5.3.4	Denosing auto-encoder(DAE)	155
5.3.5	Proposed Stacked Sparse Denosing Autoencoder(SSDAE)	156
5.4	Results and Discussion	157
5.4.1	SSDAE-based handwritten English character recognition using S-HCR dataset	157
5.4.1.1	Designing the SSDAE Network Architecture	158
5.4.2	Performance Measures	159
5.4.2.1	Analysis of the effect of different corruption rates on SSDAE	162
5.4.2.2	Ablation Study: Justification for added sparsity	164
5.4.2.3	Feature Visualization	165
5.4.3	Scalability Analysis	167
5.4.4	Comparison with state of the art methods	168
5.4.4.1	Run Time Analysis	172
	Low-Cost Device Performance Evaluation	173
5.4.4.2	Application: Exploring Potential Use-Case Scenarios	173
5.4.5	Threats to Validity	174
5.5	Summary	175
6	Enhancing Dynamic Hand gesture recognition with sEMG-Based Multi-Modal Sensors Fusion	177
6.1	Introduction	178
6.1.1	Problem statement and overview of the solution	179
6.1.1.1	Problem statement	179
6.1.1.2	Overview of the solution	180
	Deep Feature Learning Using Multimodal Deep Autoencoder (MDAE):	182
6.1.2	Major contribution of this work	183
6.2	Materials and methods	184
6.2.1	Classification of Data Fusion techniques	184
6.2.2	Dataset: Multi-Modal Handwritten Character Recognition (MM-HCR) dataset	185

6.2.3	Signal preprocessing and filters	187
6.2.4	Feature Extraction	187
6.2.5	Multi-modal Deep Autoencoder for Feature Fusion	188
6.3	Results and discussion	191
6.3.1	Multi-modal deep auto-encoder:Architecture design and implementation details	191
6.3.2	Performance Measures	193
6.3.2.1	Visualization of Learned Representations	194
6.3.2.2	Scalability and Performance Analysis	195
6.3.2.3	Time complexity Analysis	197
	Low-Cost Device Performance Evaluation	200
6.3.3	Comparison with state-of-the-art approaches	200
6.3.4	Threats to Validity	202
6.4	Summary	203
7	Improving sEMG-based dynamic Hand Gesture Recognition through Salp Swarm Algorithm-Driven Feature Selection and Fusion	207
7.1	Introduction	208
7.1.1	Problem Statement and overview of solution	210
7.1.1.1	Problem Statement	210
7.1.2	Major Contributions of the Work	213
7.2	Materials and methods	214
7.2.1	Data acquisition	214
7.2.2	Dataset	215
7.3	Preliminaries concepts and proposed method	216
7.3.1	Salp Swarm Algorithm: An introduction	216
7.3.2	Salp Swarm Algorithm for Feature Selection	217
7.3.2.1	Solution Representation	217
7.3.2.2	Fitness Function	218
7.3.3	Modified Salp Swarm Algorithm	218
7.3.3.1	Including a Weight Factor	220
7.3.3.2	Hybridizing with Particle Swarm Optimization	221
7.3.3.3	Local Search Algorithm	222
7.4	Results and Discussion	224
7.4.0.1	Experimental Setup	226
7.4.1	Multi-modal sensors feature fusion for handwritten character recognition	227
7.4.1.1	Visualization of MSSA-Selected Features using t-distribution Stochastic Neighbor Embedding (t-SNE)	229
7.4.2	Evaluating the MSSA performance on publicly available UCI dataset	231
7.4.2.1	Comparison with proposed MSSA with others published SSA modifications	234
7.4.3	State-of-the-art comparison	237

7.4.4	Threats to Validity	239
7.5	Summary	241
8	Leveraging handwriting dynamics, explainable AI and Machine learning for Alzheimer prediction	243
8.1	Introduction	244
8.1.1	Major contribution of the work	245
8.2	Methods and Materials	246
8.2.1	Dataset	246
8.2.2	Proposed Model: Stacking ensemble Classifier for Alzheimer’s Disease Detection	246
8.2.3	Shapley Additive Explanations for Alzheimer’s Disease Detection: An explainable AI approach	247
8.3	Results & Discussion	248
8.3.1	Performance Measures	248
8.3.2	Feature interpretability and ranking using SHAP	249
8.3.3	Comparison with state-of-the-art methods	252
8.3.4	Threats to Validity	253
8.4	Summmary	254
9	Conclusion and Future Work	257
A	List of Publications	261
A.1	Journal Papers	261
A.2	Conference Papers	262
A.3	Patents	263
	Bibliography	265

List of Figures

1.1	Thesis Structure	16
2.1	Myo Armband	39
2.2	A subject wearing Myo Armband during our dataset collection	39
2.3	Raw sEMG signals corresponding to six hand grasp movements used in the dataset	41
2.4	Different hand gestures for digits 0 to 9 in ASL-10.	42
2.5	Raw sEMG signals patterns of a subject for digit 9 of ASL	42
2.6	Figure illustrating the different ASL manual alphabets used in ASL-24 dataset	43
2.7	The figure illustrating the template corresponding to 26 lower cases English alphabet used for collected raw signals	44
2.8	Experimental setup used during data collection	44
2.9	A participant writing with the MyoArmband on the whiteboard.	46
2.10	The 26 lowercase English template used during data collection.	46
3.1	Generalized pipeline proposed for ASL gesture recognition task	70
3.2	Data collection setup	73
3.3	Raw sEMG signals patterns of a subject for digit 9 of ASL	73
3.4	Data collection protocol	74
3.5	Classification accuracy achieved for ten-fold cross-validation with respect to the number of iterations of classification algorithm	85
3.6	MCC score plotted against the number of Catboost iterations for each of 10 fold cross validation for ASL-10 dataset	86
3.7	The Kappa values calculated for multi-classification performed at the ASL-10 dataset	86
3.8	Figure illustrating the Multiclass loss function using 10 fold cross validation(ASL-10 dataset)	87
3.9	Multi Class AUC calculated for ten-fold cross-validation calculated for ASL-10 dataset	88
3.10	The schematic diagram describing the statistical test performed to validate the results	90
3.11	MCC score plotted against the number of Catboost iterations for each of 10 fold cross validation for ASL-24 dataset	93

3.12	Figure illustrating the Multiclass loss function using 10 fold cross validation(ASL-24 dataset)	94
3.13	The figure illustrating the time required to train various numbers of samples using LDA	95
3.14	The accuracy obtained for the Ninapro subset using the proposed pipeline	95
3.15	Features distribution corresponding to ten classes of ASL-10 dataset using t-sne	102
3.16	2D-classwise features projection using t-sne (Ninapro dataset)	102
3.17	Classwise feature importance illustrated using SHAP	103
3.18	Figure illustrating the impact of feature values on the model performance while predicting class 1 of the ASL dataset	104
4.1	Schematic diagram of the pipeline used for hand grasp recognition	112
4.2	Normalized confusion matrix archived for Channel-I	126
4.3	Normalized confusion matrix obtained for Channel-II	126
4.4	ROC AUC curve obtained for Channel-I	127
4.5	ROC AUC curve obtained for Channel-II	127
4.6	Various curves to access the scalability and performance of the two channels, Channel-I and Channel-II	132
4.7	Figure illustrating the most significant features and their impact on model output for the six classes	132
4.8	Real world application scenario utilizing IoT infrastructure	134
5.1	Generalized pipeline proposed for Handwritten character recognition task	143
5.2	The figure illustrating the template corresponding to 26 lower case English alphabet used for collected raw signals	145
5.3	The experimental set-up used for data collection	149
5.4	Alphabets writing frequency used to calculate the time taken to write a single character	149
5.5	A single hidden layer autoencoder architecture	152
5.6	Stacked autoencoder structure	153
5.7	ROC AUC curve obtained for SSDAE + SVM on S-HCR dataset	161
5.8	Precision, Recall, and F1 Score obtained for SSDAE + SVM on S-HCR dataset	161
5.9	Classification accuracy achieved with varying corruption rates for SSDAE and SDAE on S-HCR dataset	166
5.10	t-sne plot showing the feature distribution learned using SSDAE network for S-HCR dataset	166
5.11	Learning curve showing training and cross-validation scores for S-HCR dataset using (SSDAE + SVM) model	169
5.12	Plot showing the model training time with respect to the number of training samples for S-HCR dataset using (SSDAE + SVM) model	169

5.13	Graph showing the relationship between the model training time and cross-validation score for S-HCR dataset using (SSDAE + SVM) model	169
5.14	A diagram illustrating a real-world smart classroom scenario where the proposed handwriting recognition model can be deployed.	174
6.1	A schematic diagram illustrating the proposed pipeline for whiteboard-based handwritten character recognition task	184
6.2	Data collection protocol	186
6.3	Raw and filtered accelerometer signal using wavelet denoising	187
6.4	Raw and filtered sEMG signals using Butterworth filter	188
6.5	Demonstration of multi-modal deep auto-encoder architecture	190
6.6	Precision, Recall, and F1 Score obtained for MMDAE + SVM on MM-HCR dataset	195
6.7	Plot illustrating the ROC AUC curve obtained for 26 classes using SVM classifier	197
6.8	Training and validation curve obtained using SVM classifier	198
6.9	Plot illustrating the scalability of the model with respect to training samples and time delay (in unit time)	198
6.10	Plot illustrating the performance of the model with respect to time delay (in unit time)	198
6.11	Plot showing the shared features learned using proposed deep multi-model autoencoder when visualized using t-sne	199
6.12	Experimental setup used to deploy the model on low-cost devices (Raspberry Pi)	200
7.1	Schematic diagram of the pipeline used for handwritten characters recognition	214
7.2	Plot of w_1 and w_2 with iterations.	221
7.3	Workflow of the proposed MSSA feature selection algorithm	223
7.4	t-SNE plot showing the feature distribution learned for dataset $MM - HCR_{EMG}$	230
7.5	t-SNE plot showing the feature distribution learned for the dataset. $MM - HCR_{IMU}$	230
8.1	Schematic diagram showing proposed stack ensemble model	247
8.2	ROC AUC curve generated using machine learning ensemble	249
8.3	Feature importance using SHAP	251
8.4	Summary plot drawn using SHAP	252

List of Tables

2.1	Summary of the related work for sEMG sensor based ASL recognition	23
2.2	Summary of the related work for sEMG sensors-based handwriting recognition tasks	27
2.3	Summary of the related work for multimodal sensors-based handwriting recognition tasks	31
2.4	Summary of recent methods focused on the minimal number of sEMG channels for classifying different hand grasp movements.	36
2.5	List of Tasks Performed and Their Categories	47
2.6	Features and their Descriptions	47
2.7	Publicly available dataset used for the experiments	48
3.1	Different performance metrics achieved using the baseline classifiers	87
3.2	The selected features and the corresponding p-values	91
3.3	Summary of the related work for sEMG sensor-based ASL recognition	98
3.4	Various features and the corresponding performance metrics	100
3.5	Response time of the proposed pipeline to identify a single ASL gesture	105
4.1	Mean Accuracy and MCC scores achieved using baseline classifiers	125
4.2	Performance metrics for baseline classifiers for Channel-I	126
4.3	Performance metrics for baseline classifiers Channel-II	126
4.4	Summary of state-of-the-art methods for classifying different hand grasp movements.	130
5.1	Proposed SSDAE parameters and Layer Details	159
5.2	The classification accuracy and mean MCC score achieved using SSDAE network and baseline algorithms	162
5.3	Precision, Recall, and F1 score achieved using SSDAE + SVM model on S-HCR dataset	162
5.4	The classification accuracy and mean MCC score achieved using SDAE network and baseline algorithms	167
5.5	Comparison of the proposed method for handwritten character recognition with state-of-art approaches	172
5.6	Training time of the different model used for HCR task	173
5.7	Response time of the proposed pipeline to identify a single handwritten lowercase English alphabet using SSDAE+SVM model on S-HCR dataset	173

6.1	Summary of the Multi-Modal Autoencoder Architecture with Parameters	193
6.2	Different performance metrics achieved using the baseline classifiers .	196
6.3	Confusion matrix obtained for 26 classes using SVM classifier	196
6.4	The average recognition time while predicting a single whiteboard-based handwritten character	200
6.5	Summary of the related work for multimodal sensors-based handwriting recognition tasks	202
7.1	Comparison of Feature Selection and Accuracy	228
7.2	Publicly available dataset used for the experiments	228
7.3	Performance comparison(Classification Accuracy) of MSSA with other methods on benchmark datasets	233
7.4	Performance comparison (Average Feature Selection Size) of MSSA with other methods on benchmark datasets	234
7.5	Performance comparison (Average Execution Time) of MSSA with other methods on benchmark datasets	234
7.6	Performance comparison (Fitness Values on 20 runs) of MSSA with other methods on benchmark datasets	235
7.7	The accuracy achieved on UCI data-set through different SSA modifications	237
7.8	Comparison with state-of-the-art approaches	239
8.1	Performance metrics achieved for our proposed stacking ensemble model	248
8.2	Comparison of the state of the art methods on DARWIN dataset . . .	253

Abbreviations

1D-CNN	One-Dimensional Convolutional Neural Network
3D	Three Dimensional
AD	Alzheimer’s Disease
AI	Artificial Intelligence
ANN	Artificial Neural Network
ASL	American Sign Language
AUC-ROC	Area Under the Curve - Receiver Operating Characteristic
AUC	Area Under the Curve
BGRU	Bidirectional Gated Recurrent Unit
BiLSTM	Bidirectional Long Short-Term Memory
BLSTM	Bidirectional Long Short-Term Memory
CSV	Comma-Separated Values
CNN	Convolutional Neural Network
DAE	Denoising AutoEncoder
DARWIN	Dataset for Alzheimer’s Research on Writing kINematics
DCNN	Deep Convolutional Neural Network
DTW	Dynamic Time Warping
EMG	ElectroMyoGraphy
FFT	Fast Fourier Transform
FPR	False Positive Rate
GRU	Gated Recurrent Unit
HCI	Human-Computer Interaction

HCR	H andwritten C haracter R ecognition
HGR	H and G rasp R ecognition
HMM	H idden M arkov M odel
IMU	I nertial M easurement U nit
IoT	I nternet of T hings
KNN	K - N earest N eighbors
KRE	K raskov E ntropy
LDA	L inear D iscriminant A nalysis
LSTM	L ong S hort- T erm M emory
MCC	M atthews C orrelation C oefficient
MDAE	M ultimodal D eep A uto E ncoder
MDC	M odel of D ynamic C ontraction
ML	M achine L earning
MSE	M ean S quared E rror
MSC	M odel of S tatic C ontraction
MU	M otor U nit
MUAP	M otor U nit A ction P otential
MSSA	M odified S alp S warm A lgorithm
PCA	P rincipal C omponent A nalysis
PSO	P article S warm O ptimization
ReLU	R ectified L inear U nit
RF	R andom F orest
RNN	R ecurrent N eural N etwork
ROC-AUC	R eciever O perating C haracteristic - A rea U nder the C urve
ROC	R eciever O perating C haracteristic
SHAP	S hapley A dditive e x P lanations
SLRS	S ign L anguage R ecognition S ystem
SSA	S alp S warm A lgorithm
SSAE	S tacked S pase A uto E ncoder

SSDAE	S tacked S parse D enoising A uto E ncoder
sEMG	surface E lectro M yo G raphy
SVD	S ingular V alue D ecomposition
SVM	S upport V ector M achine
TQWT	T unable- Q W avelet T ransform
TPR	T rue P ositive R ate
t-SNE	t - D istributed S tochastic N eighbor E mbedding
TP	T rue P ositive
UWB	U ltra- W ide B and
XAI	E xplainable A rtificial I ntelligence
XGB	X G B oost

Symbols

D	Dataset consisting of sEMG time series recordings
N	Number of sEMG time series recordings
$TS_i(u)$	Time series recording i with data points u_1, u_2, \dots, u_N
$f_{i,m}$	Feature extracted for a window size where m is the total number of features
Ins_i	i^{th} instance in the dataset
l_k	Class label of hand grasp movement corresponding to the instance Ins_i
$v(S)$	Real-valued profit/payoff for coalition S
$\delta(S, j)$	Marginal contribution of player j to coalition S
$\varphi_j(X, v)$	Shapley value for player j in the coalition game
$G(X_d(t_n), t_n)$	Deterministic part of the Langevin equation
$H(X_d(t_n), t_n)$	Stochastic part of the Langevin equation
$\tau(t_n)$	White Gaussian noise in the Langevin equation
X	Raw sEMG data matrix
T	Number of time samples
$\Phi(X)$	Feature vector generated from raw data
M	Number of features
$\Phi_{S_i}(X)$	Set of relevant features selected by algorithm i
$\Phi_S(X)$	Combined feature vector from multiple methods
h	Classifier function
θ	Parameters of the classifier
\hat{Y}	Predicted gesture label

Symbols

Y	True gesture label
$L(\theta, S)$	Classification loss
ℓ	Loss function (e.g., cross-entropy loss)
A	Classification accuracy
A_{\min}	Specified accuracy threshold
$T_s(u)$	Time series corresponding to each sEMG sensor
m	Data points in time series
$W_s(p, s)$	Window segment of length p
z	Total number of samples
l_i	Annotation (label)