**Short thesis for the degree of doctor of philosophy (PhD)**

# Prediction and Analysis Based on Sensor Network Data Using Machine Learning Techniques

by Ashraf Khaled Abd Elkareem Aldabbas

Supervisor: Dr. Zoltán Gál


UNIVERSITY OF DEBRECEN
Doctoral School of Informatics

Debrecen, 2021

# Outline

# 1. Introduction

The incorporation of planetary science and environmental events in the ML method is a unique challenge. Spacecraft functioning in deep space will produce vast data amounts because of its mission complexity. Exploiting ML in planetary missions varies from other prevalent implementations. The essential use of the acquired planetary data is to evolve the fundamental scientific hypotheses.

Recently, interest in science and engineering focused on Machine Learning (ML) has increased rapidly. This growing excitement stems from the collaborative development and use of successful computational algorithms, large quantities of data accessible from experimental tools and other outlets, and achievements reported by researchers and academics. Contemporary developments in the data obtained by outer space expeditions have provided the space required for groundbreaking study and classification models. Integrating science domain expertise can significantly minimize data requirements and improve testing and prediction. Spacecraft functioning in deep space will produce vast data amounts because of its mission complexity and various orbits and rotations that can be made. Conventionally planetary branch of knowledge has utilized gist scientific techniques like remote sensing.

Along with the contemporary attainability of sampled circumference data offered by such outer missions, the ML method provide remarkable capability preference. Exploiting ML in planetary missions varies from other prevalent implementations. Cassini's data generated by the orbiting Saturn system combines temporal and spatial extraordinary observations. This engagement requires (Spatio-temporal) complex events detection and in-depth analysis [1]. Complexity relies on what comprises the general scope of knowledge. The scientist's role is to illustrate various models based on their field. This brings about a model that should be relatively unpretentious to be comprehended by their prediction procedure; the algorithm must bring forth an optimal explicable model for effectively dealing with computationally tricky issues. Significantly, the concepts

regarding performance [2]. This work is interested in analyzing the interplanetary and environmental domain to acquire scientifically insightful outcomes from the enforced ML models. The twofold defiance of Spatio-temporal generated data of spacecraft orbiting planets are constituted for the spacecraft data transmission and observation. A variety of data sizes establishing an approach to classify special and complex events within the Cassini dataset. Just as Saturn's distinctive data set and the ambiance of its surroundings, which provides exquisite perception toward applying detection and classification models into interplanetary data via ML, to acquire higher accuracy and interpretability of detected complex events.

In this thesis, structured ML prediction is considered and precisely, those involving sequential structure. It is focused on the relatively mature methods of complex event processing, which is associated with the identification of complex events focused on domain specialist rules and trends, while the meta-learning LSTM and Constructive Knowledge-based Event (CKE) algorithms can provide better analyses of vast volumes of data within a small-time interval and supervised learning via structured machine-learning prediction. This dissertation focus on the specific issues prompted by structured outputs when the object involved in a ML task has a complex event processing, prediction, or multi-class classification where the goal is to predict several outcomes from some set to an instance.

I present my structured ML approaches to learn specific similarity measures for change-point detection. Analyzing the interplanetary trajectory is a big part of Saturn’s task research as rapidly linked tools, and sensory devices become part of our everyday lives. The high-velocity knowledge flow sea is rising. This vast amount of high-rate data generated demands rapid insight in numerous applications such as the Internet of Things, energy storage, etc. This produces the need for Complex Event Processing (CEP) frameworks, utilizing articulate state-of-the-art approaches to collect qualitative details. The Mission of Cassini, as an illustration, generated more than 630 gigabytes of research-based data that include 450,000 taken images. ML assists the experts and researchers with data enclosed by this vast extent. This dissertation uses the Cassini dataset as a particular instance of analysis to illustrate the remarkable capabilities of introducing Artificial

Intelligence (AI) among space missions to extend further intelligent computing. It is intended to consider exploiting Deep Learning (DL) on the space missions evolution platforms, offering higher efficiency and reliability. Using DL classifiers with diverse data volume access, it is illustrated that incorporating the collected spacecraft data with machine-learning approaches, which is fundamental for obtaining scientific significance. Based on these findings, the provided models on incorporating space sensed data into AI scope earmarking supervised classification concerning planetary data, which expresses a path incorporating Cassini spacecraft mission data into ML. The techniques of DL can be utilized to evolve intelligent solutions. Over the past several years, massive progress adoption of AI techniques with special attention on neural networks has appeared globally. There has been a generous dash in the AI scope to provide robust solutions in numerous fields. Models focused on Recurrent Neural Network (RNN) with Long Short-Term Memory (LSTM), Bidirectional LSTMs, and Gated Recurrent Unite (GRU) are exceptional in learning sequences and capable of capturing long-range dependencies in the temporal data collection. In this analysis, various LSTM and GRU models are used to model the Cassini–Huygens' outer space mission. The "learning" particular aspect of ML indicates an algorithm's potency to detect data patterns to enhance the results, i.e., to utilize the available data to inspect and predict the unrenowned. ML has many implementations in video surveillance, online banking, aviation, product recommendations, and so on. Moreover, the technology is anticipated to leverage the forthcoming space exploration since it can process massive data volumes, foretell spacecraft status, and detect patterns in the analyzed images. ML could empower the cost-effectiveness, science profit, and dependability of missions in outer space.

This work employs Cassini's planetary mission data and environmental events to build DL classifiers. It is also demonstrated that amalgamating the generated spacecraft data ameliorates the performance and delineation of ML approaches, which is fundamental for obtaining systematic purport. Based on these accumulations of findings, an approach amalgamating space generated data into ML methods designating supervised classification within the scope of planetary data context. These reached results provide a

footpath for incorporating data of planetary missions and DL algorithms to grant the computers the ability to impart data to create categorize and predictions swiftly and with high accuracy. A cutting-edge predictive CEP system has been established that uses historical knowledge to diagnose unique and complicated incidents. To use historical knowledge effectively, the method uses N-dimensional, historically matched sequence space. The prediction may then be made by addressing the set of queries over historical sequence space.

The main publications related to this work are [C1, C2, C3, C4, C5, C7, J1, J2, J3, J4, J5, J6], the other publications include [J7, C6, C8, C9].

## 2. Problem Statement and Motivation (Research Aim)

Science research focuses on two pillars: data collection, which utilizes computations, tests, and/or observations to produce new measures of complicated phenomena; and data processing, which seeks to derive new information from the data. Historically, much of the work has been to gather evidence, for example, beginning with the push to build telescopes to study planets and finishing with large-scale, multi-physics simulations to obtain a glimpse into potentially inaccessible processes, such as supernovae. As data size and sophistication grow, the risk of missed breakthrough opportunities may also intensify and will potentially impede development dramatically. To overcome this problem, several research fields shift. In the evolution procedure of complex systems, interdisciplinary teams of scientists and engineers are engaged. Frequently, such teams are established by inter-institutional collaboration.

Thus, the training set must be broad enough to construct a generic model from it, which performs well with the test process's unseen data. Training such models will cost computationally. However, once qualified, they work quickly in the implementation process. However, thanks to fast CPUs and GPUs' availability, computing costs are no longer a concern. Different academic papers show a concrete proof of ML algorithms [3-4] success in pattern detection, time series estimation, data clustering, and classification.

In this research analysis, outer space mission modeling is conceived as an issue of time-series prediction (temporal sequence modeling), so the ML algorithm can be implemented. In such a formulation, occurrences and their complexity function as triggers and, therefore, collectively determine the device's state, and sensor values are the resulting consequences. More data-driven approaches eventually seek to eliminate the need for assumptions and replace them with meticulously tailored theories with extensive data collections. This data explosion is not restricted to research applications. Investments in large data from science missions, tools for predictive modeling and algorithms, high-performance computing systems would allow breakthroughs and significant change. Incorporating a variety of Deep Learning (DL) models, from the unpretentious to the more sophisticated, in addition to differing datasets, will provide insights into the essence of the data produced by the underlying spacecraft-generated data. DL applications in space science and big data analytics are both the high focal point of data science. This set of methods has been confirmed to be competent to swift and enhance both essential and empirical research. I implement a comprehensive overview and examination of the most up-to-date research in this scope.

For the time being, ML algorithms are effectively utilized for prediction, dimensionality reduction, and classification tasks of big datasets [5]. ML has been evidenced to have exceptional capacity in great domains such as image classification [6], speech recognition [7], self-driving cars [8], and many others.

## 3. Methodology

CEP is a meta-framework of techniques, e.g., filtering events, matching event patterns, timing analysis, a hierarchical abstraction of events, creation of complex events, the definition of event hierarchies) for processing event flows in real-time and abstracting humanly understandable and actionable knowledge from such event flows. By comparison, AI has experienced

colossal development in its potential definitions from its initial inception. The summary of this work is divided into six main parts discussed below:

The presented theses outline the experimental setups used to test the research hypothesis. Sensory data are explained, and their dependencies are identified with each other. The theses realize the methodology to satisfy the theories. It discusses the adopted work methodology. It presents and describes numerous phases in the execution of the methodology aspect. It illuminates the preparation of training data and shows the neural network's design and other technical information.

The first thesis is concerned with adopting remote sensing as a tool to detect climate comfort and modeling the events as 3D+ spaces (having at least three dimensions, such as spatial and temporal). For each sensor, one may determine qualitative events dependent on sensor measurements (e.g., temperature, humidity) and different degrees of granularity for each merit (e.g., day-month-year for the period).

The second thesis propose method identified a collection of complex events using the weighted Complex Event Level (wCEL) metric across five evaluated phases of the Cassini project, namely the Approach science, the Extended mission, the Extended-Extended mission, the Tour mission, and the Tour pre-Huygens trip. In the analysis process, fifteen metadata variables were included.

The third thesis provides the element of learning and reasoning with structured prediction based on revealing event complexity is given, whereas the task of revealing complex events encompasses modeling special events that are structurally associated. The put forward algorithm is intended to learn the prospective functions of reasoning with structured prediction based on revealing event complexity, an amalgamation of outstanding conceptual nonlinear features conveyed by regression models.

In the fourth thesis, the Constructive Knowledge-based Event (CKE) algorithm is introduced, which provides associate degree creative methodology of CEP to fully develop process patterns or events counting on the constructive feature computing. I presented a constructive event detection technique to find situations when the event’s statute has

transferred over a special event to a complex event by generating an emphasis centerpiece on complex interconnection components by utilizing stacked bidirectional Long Short-Term Memory (LSTM) networks.

The fifth thesis presents the Meta-Learner LSTM algorithm; it speculates an essential topic related to the classification of remotely sensed images, in which the process of feature coding and extraction are decisive procedures.

The sixth thesis allocates a sophisticated in-depth learning approach for detecting Cassini spacecraft trajectory modifications in post-processing mode. The model utilizes the ability of Recurrent Neural Networks (RNNs) for drawing out useful data and learning the time series inner data pattern. The proposed approaches cast the learning as a structured prediction, one with losses adequately designed for the task: the F1 and Matthews correlation coefficient. In the following section, I will discuss the results achieved in this work and refer each to its appropriate publication.

## 4. Simulation Findings and New Results

This part presents the assessment results of generic environmental data analysis and several RNN model instances produced during this research. Moreover, model extensibility in space is evaluated. It starts by reporting findings and evaluating all models offered. Evaluation results of several RNN models established in this analysis are provided. This section contributes supervised data training for timestamped data on which the previously provided section results are given. It is not a straightforward task to extract information from an interplanetary spacecraft mission. Alignment utilizes dissimilarity measures among timestamped data. Then supervised learning models are proposed, accompanied by a threshold scale. The detection approach is also cast as a prediction task using the F1 score and the Matthews correlation coefficient (MCC). It demonstrates that it is feasible to the score, particularly to include meaningful results and interpretation. Ultimately, experiments on a real dataset are provided. The coverage of the hypotheses outlined in the first chapter is presented in the last part. Since it is not enough to propose a new algorithm, several different

computational tests are performed to verify the algorithm's functionality to ensure that the proposed algorithm is efficient. A decision (classification), or a forecast, is the outcome of an ML process. Scientific computing has historically been dominated by complex numerical simulations that are intensive in terms of resources. However, new smart possibilities are generated with the advent of data-driven models and algorithms.

## 4.1. Generic Environmental Models

Since it is not enough to propose a new algorithm, several different computational tests are performed to verify the functionality of the algorithm to ensure that the proposed algorithm is efficient. A decision (classification), or a forecast, is the outcome of an ML process. I need the opportunity to calculate ML efficiency rigorously for accurate and credible usage of ML in these results.

Thesis 1 (J6, C7)1 (Section 5.1)2 examines the use of remote sensing to detect climate comfort and the simulation of incidents as 3D+ spaces (having dimensions, such as spatial and temporal or CO2 emission level). Day-month-year, each parameter is often assigned for each event, and granularity is highly relevant: (qualitative): e.g., on the premise that day-month, degree of the granularity depending on measurement values are calculated for each sensor's measurements (or event).

### 4.1.1. Healthy Residential District Detection Approach Results

As I do not have a timestamp for the collected data, by performing a splitting of data into a smaller range, the storing order of the 2% of the samples gave us the road formation trajectories. This information made us consider that the sampling order matches the data's storing order in the file. The interpretation process provides visual depictions that include a GIS view utilized to produce the spatial contour plot representation. This provides visual relationships between the 3D coordinates of the samplings and the CO2 emission levels through a parallel contour view. These figures

are sophisticated representations of environmental sensory data sets showing the high number of measurement points, drawing with good accuracy even the map of the peninsula (see Figure 4.1). In our model, the yellow contour lines represent a higher emission level, which means more polluted zones, while the blue points represent a less polluted zone and the purple is the least, which are the healthiest residential district to live in. If I compare the contour plot in Figure 4.2 with Google map view Figure 4.1.

*Table 4. 1: CO2 levels and categories*

| $CO_2$ level categories | $CO_2$ level regions |
|---|---|
| Most polluted zone | $R_{10,6}$, $R_{9,6}$, $R_{6,7}$, $R_{6,8}$, $R_{5,7}$ |
| Reasonable zone | $R_{3,7}$, $R_{3,8}$, $R_{9,1}$, $R_{9,2}$, $R_{9,3}$,$R_{10,6}$,$R_{2,6}$,$R_{2,7}$,$R_{2,8}$ |
| Healthy zone | $R_{9,7}$, $R_{7,2}$, $R_{8,2}$, $R_{10,3}$, $R_{1,8}$, $R_{1,9}$ |

While the most convenient regions are: R9,7, R7,2, R8,2, R10,3, R1,8, R1,9. Which can be organized as a table below:

*Figure 4. 1: Google map view of the Peninsula*

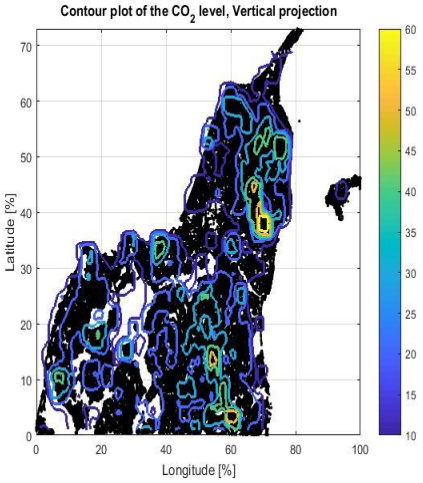


*Figure 4. 2: Contour plot of the $CO_2$ level from vertical projection*

[1] "J" denotes Journal papers, "C" denotes Conference papers
[2] Here, the related sections of the dissertation are referred to.

Subsequently it can be concluded that the most affected regions with high CO2 emission levels are: R10,6, R9,6, R6,7, R6,8, R5,7, followed by the regions of the green color: R3,7, R3,8, R9,1, R9,2, R9,3, R10,6, then followed with the regions of the blue such as the regions R2,6, R2,7, R2,8, as an example.

### 4.1.2. Climate Comfort Detection Approach

When a historical dataset is acquired and investigated, it usually obeys a particular pattern known as statistical distribution, as an example. If (116) years of January precipitation data were gathered and examined, the pattern of the data has which composed of the January precipitation as an example having a trend to a very slight decrease that can be concluded by the value of -0.066 over the mentioned period of time, as illustrated in Figure 4.3.

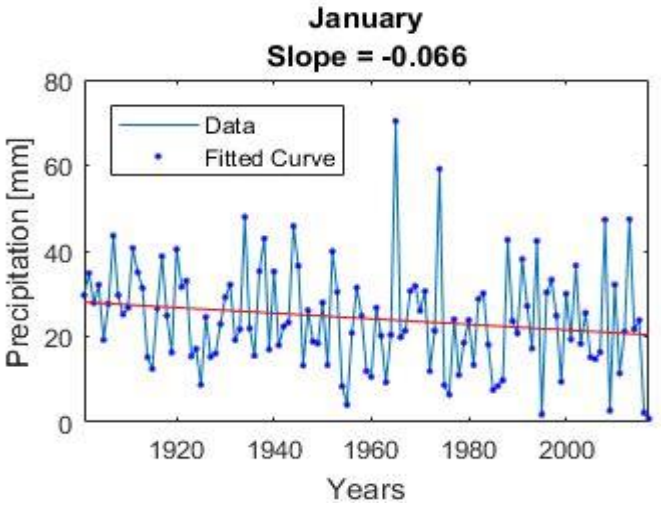


*Figure 4. 3: Precipitation-January slope over 116 years*
*(Data: line of segments; Fitted curve: dots; Trend: red line)*

It can be concluded that the average annual precipitation is around 15 inches, and the average annual temperature is around 10.5 °C. It is significant to appraise how climate holds numerous and varied data. The

mean monthly precipitation and temperature could be sketched to represent seasonality by month. The next chart illustrates the mean monthly precipitation and temperature for the Jordan region through the period (1901-2017). While the average temperature for July shows a slight increase over that period with a slope value equals to 0.020, which can be seen in Figure 4.4.

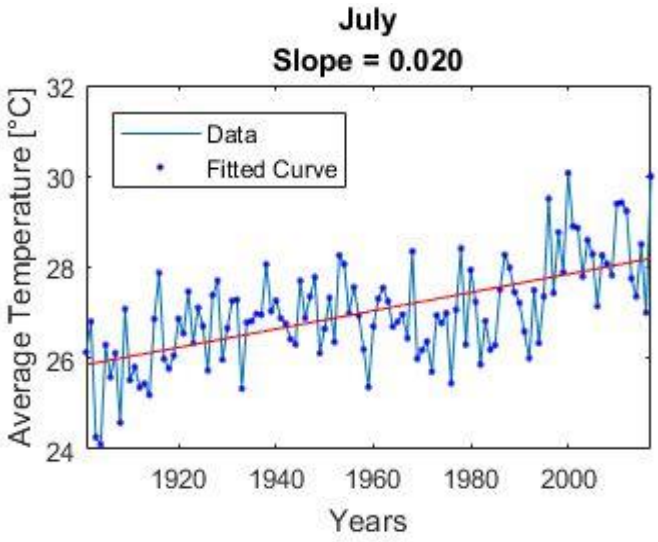


*Figure 4. 4: Average temperature-July slope over 116 years (Data: line of segments; Fitted curve: dots; Trend: red line)*

To show the modification precipitation and temperature in a monthly manner, gridded monthly precipitation/temperature time-series data obtained and utilized to provide a mean monthly/annual climatologies version archive, the obtained monthly modification of precipitation and temperature (1901-2017), is shown in Figure 4.5 this time series illustrates that there is a slight change over the given period.

The generated Figure 4.6 shows the TCI value of the Jordan area for each month. The scores of TCI are slightly greater in several months of autumn

and winter (March, April, and October) rather than other months, correspond to the categorization of TCI distributions. The prevailing seasonal modality in Jordan is the peak of the winter as it can be seen in Figure 4.6 the maximum TCI score is in (April, followed by March, May, and October), while in the summer period (June, July, August, and September), the scores are proportionately weak, so that coincides with the category "Marginal". July is observed as the worst month to conduct any activity for tourism in terms of climate. TCI's five sub-indices participate variously to the score of TCI for every month, the sub-indices vary from month to month, and the contrasting climatic sturdiness. The maximum TCI score is more than (80) in April, consequently categorizing as having 'excellent'. Other months such as March, May, and October are in the classification of 'very good'.

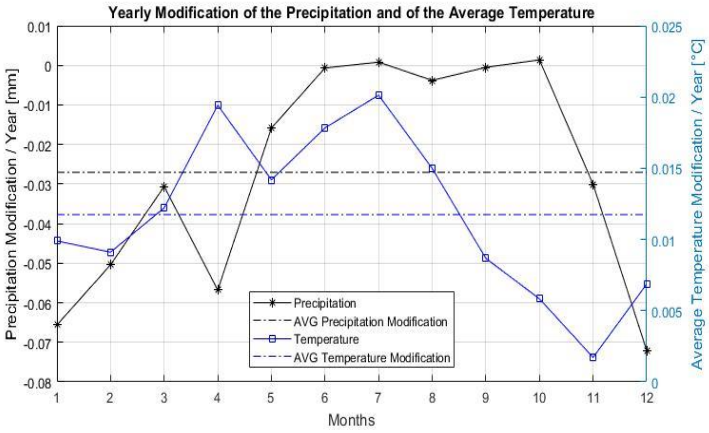


*Figure 4. 5: Monthly modification of precipitation and temperature (1901-2017)*

October is the least among them. January, November, and December are considered among the classification of 'good '. It can be noticed that their related scores are nearby to each other. There is an indication to mention that higher summer temperatures take part in reduced tourism

disbursements in Jordan. It is clear that June, July, August also September possess the greatest hits of temperatures, and this will affect their attendance by having inferior circumstances for tourists. Sunshine and precipitation participate to enlargement of TCI score, wind speed owns the lowest values in the months of the summer, particularly in July and August.

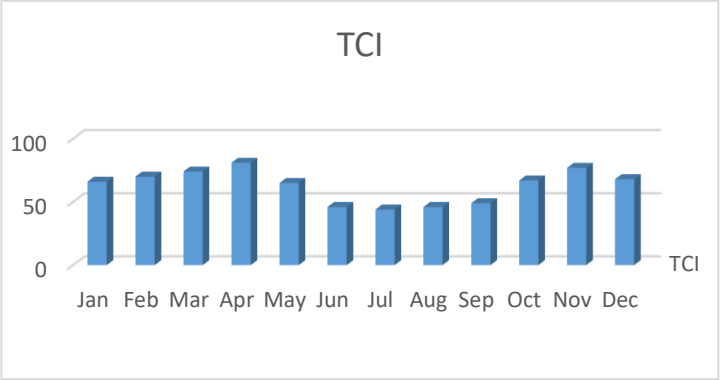


*Figure 4. 6: Monthly TCI in Jordan (2011-2018)*

Thus, it must be observed that the wind speed to the greatest extent is the undesirable element, as it decreases the score of TCI in summer due to snug wind also in January due to wintry wind.

## 4.2. Detected Complex Events of the Cassini-Huygens Mission SED and wCEL variables evaluation

The proposed method has detected a list of complex events through the metric of weighted Complex Event Level (wCEL), among Cassini's image batches, among five analyzed phases of the mission, which include: Approach science mission, Extended mission, Extended-Extended mission,

Tour mission, Tour pre-Huygens mission. Fifteen variables of the metadata were involved in the analysis process.

Thesis 2 (C1, C2) (Section 5.3) An innovative method is introduced in this research to detect complex events by analyzing a massive dataset that has been collected using the remote sensing approach. The proposed method is influenced by the indexed time-series manner using a score array and the variation score of the target time intervals via the wCEL, allowing to convert the sensed data into spectacular detected events. The results showed that there was more than a single complex event in almost every batch.

Transmission of the samples to the Earth was influenced by Saturn, Cassini, and Earth's relative position causing variable temporal storage durations of the spacecraft samples. Mean weighted complex event level with different window sizes for the group (A) is given in Figure 4.7. While the coefficient of variance is provided in Figure 4.8. Mean weighted is correlated with an occurrence or result with its consistent quantitative outcome and then outlining all entities together.

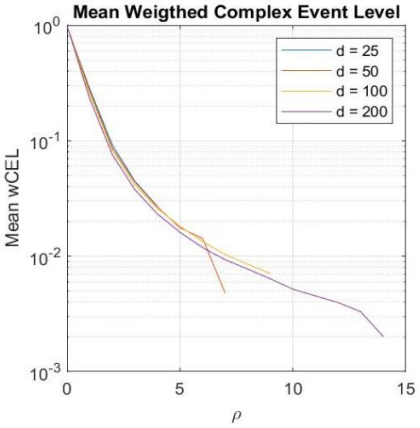


*Figure 4.7: Mean wCEL (Group A) for different parameters, d (sampling window sizes)*

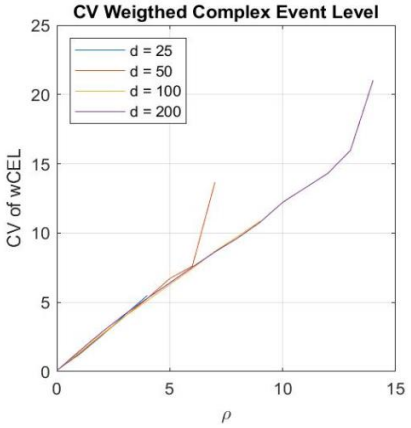


*Figure 4.8: Coefficient of Variance of wCEL (Group A)*

Based on Figure 4-9, it can be concluded the cut values of the complex event-level detector mechanism for group A of the variables (see Table 4.1).

*Table 4.1: The estimated cut value of the sensitivity of the average wCEL (Group A: timing)*

| **Sample window size (d)** | $\boldsymbol{\rho^*}$ |
|---|---|
| 25 | 5 |
| 50 | 8 |
| 100 | 10 |
| 200 | 15 |

Figure 4.9 represents the mean wCEL (Group A) estimation of timing, sampling window, d = 200). It was found that the mean weighted Complex Event Level metric for the timing is an exponential function of an exponential function and has the following formula:

$$mwCEL(\rho) = \begin{cases} 1 & if\ \rho = 0 \\ \exp(-a \cdot \rho \cdot \exp(-b \cdot \rho)) & if\ \rho \in (0, \rho^*) \\ 0 & if\ \rho \geq \rho^* \end{cases} \quad (4\text{-}1)$$

where $mwCEL$ is the mean of weighted Complex Event Level, $\rho$ is the sensitivity of the Special Event Detector, $a$ and $b$ are parameters depending weakly on the sample window size (d). Parameter $\rho^*$ is the cut value of the SED sensitivity and based on Figure 4.9 depends on the sample window size: the greater is $d$, the greater becomes $\rho^*$. This phenomenon is caused by the smoothed property of the SED influenced by the moving average function.

Estimated fitting parameters $a$ and $b$ are given in Table 4.2 with the coefficient of determination, $R^2 > 0.99$.

*Table 4.2: Estimated parameters of the average wCEL (Group A: timing)*

| Sample window size (d) | *a* | *b* | $R^2$ |
|---|---|---|---|
| 25 | 1.353 | 0.070 | 0.999 |
| 50 | 1.413 | 0.086 | 0.999 |
| 100 | 1.483 | 0.085 | 0.999 |
| 200 | 1.594 | 0.103 | 0.999 |

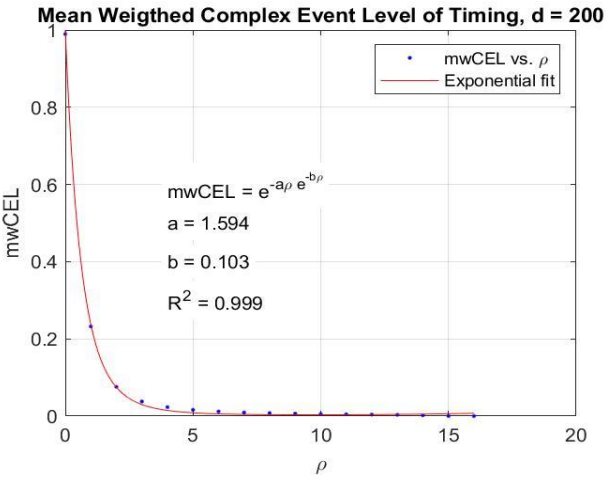


*Figure 4.9: Estimation of the mean wCEL (Group A) of timing, sampling window, d = 200)*

## 4.3. Structured Prediction Based on Revealing Event Complexity

Sensor technology is essential to advance analytics. Sensors generate data signatures requiring more off-line analysis. The ability to detect variance within incoming data is an essential track of this model.

Thesis 3 (J1) (Section 5.2) A state of art speculative is provided with the capacity to gain an accurate and deep intuitive understanding of structured prediction. The model gaze to immense boundary speculation in a vast domain of prediction patterns wheresoever reasoning encompasses puzzle out complementary idealization. The put forward algorithm is intended to learn the prospective functions of reasoning with structured prediction based on revealing event complexity.

A robust method for identifying variation is utilized to detect if an event has occurred in a particular image time, based on an assessment made with the taken image timing variables, which include image meantime; capture start/end time of the sample is illustrated in Table 4.3.

*Table 4.3: Inspected mission phases*

| Phase Name | Start Sample ID | Start Time | End Sample ID | End Time |
|---|---|---|---|---|
| APPROACH SCIENCE | 1 | 2004-037T02:07:06.498 | 10,675 | 2004-162T14:47:05.854 |
| TOUR PRE-HUYGENS | 10,940 | 2004-164T02:33:51.000 | 31,640 | 2004-358T13:47:22.548 |
| TOUR | 32,032 | 2005-015T18:28:29.491 | 166,187 | 2008-183T09:17:06.323 |
| EXTENDED MISSION | 166,188 | 2008-183T21:04:09.008 | 235,887 | 2010-283T14:14:20.741 |
| EXTENDED-EXTENDED MISSION | 235,888 | 2010-285T05:23:32.745 | 407,303 | 2017-257T19:59:04.075 |

Within the analysis, I utilized a three-valued array [X, Y, Z] components of the spacecraft position vector in [km]. Figure 4.10 shows the position in the sampling range: 45000-65000. Timing is connected with every spacecraft locomotion information in an observation. Timing enumerates the whole image events in each stamp and associated start and end times. The aspect of “learning” with ML indicates the aptitude of an algorithm to notice and learn patterns among data to ameliorate the results, i.e., to utilize the existing data to foretell events and solve the ambiguity. I’m exploiting the LSTM network as our adopted framework. The LSTM framework input is a sequence such as a sample number, sampling moments, and observations

among the intervals. After executing the previous sequencing of temporal modeling, I hand over the output to LSTM.

The output is passed into two fully connected layers, these layers are required for generating an eminent representation which facilely provides a discriminate, as closure the LSTM framework is fed to an output regression layer that comprises time steps, each event represents a sample observation that may bring about a complex event.

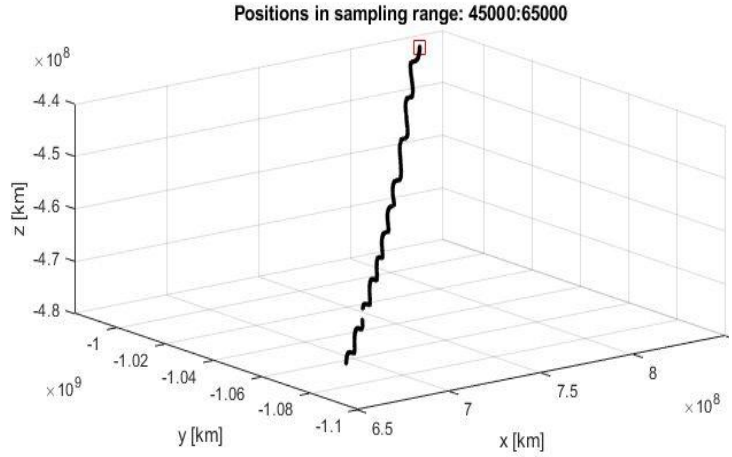


*Figure 4.10: Cassini's [x, y, z] Position*

Also, the observed variables will represent the features of the samples. Figure 4.11 represents the velocity magnitude vector change and its histogram; based on this graph, it was found that the number of velocity magnitude changes during the (2004-2017) Cassini-Huygens project was (~ 10k). Time steps offer an additional method to express our time series; the algorithm considers the previous time series to be as a foretell for the following time step. The model has a total number of layers, which equals 5, the input layer of LSTM has 100 hidden units, while the parameters of the fully connected Layers for numClasses1 = 100, and numClasses2 = 1.

LSTM cell has another number of parameters to train our model via Adam optimizer over the max epoch, which equals 100 epochs with minibatch size = 128 and validation frequency = 10 of bracing the performance. There are two fully connected layers. The model ends up with a regression-layer to get the final result. To assess the efficiency of our ML approaches for actuating complex events to be specified among large datasets, I have applied the model to reveal the complex events related to velocity magnitude change.

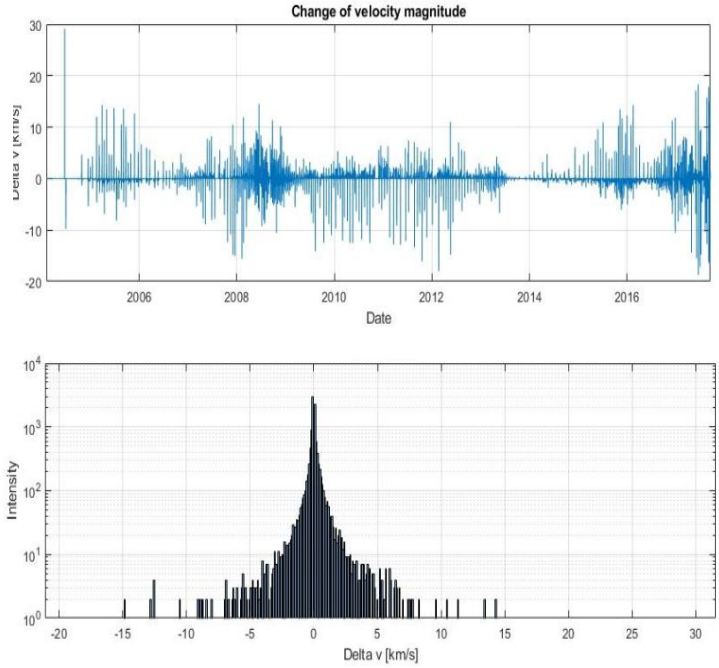


*Figure 4.11: Velocity magnitude change and related histogram over time*

The histogram velocity plotted in the logarithmic scale shows the velocity magnitude change over time. The algorithm was able to reveal the related complex events. The main aim behind applying our classifier is to specify

complex events from sensory generated data, detect temporal semantics for complex events recognition, and reveal them.

## 4.4. Constructive Knowledge Based Event Model

Thesis 4 (J2, C3) (Section 5.4) I present a constructive event detection technique to find situations when the event's statute has transferred over a special event to a complex event by generating an emphasis centerpiece on complex interconnection components by utilizing stacked bidirectional Long Short-Term Memory (LSTM) networks.

The Constructive Knowledge-based Event (CKE) detection method tested on the data set provides over (90%) of the Saturn/Cassini-Huygens interplanetary project's special events hit rate. This approach empowers analyses of vast volumes of data within a small-time interval.

The used dataset is downloaded from NASA's jet propulsion laboratory. The part is concerned with the data captured during the period that extends between (16-Feb-2004, 15-Sep-2017), which is almost 13.5 years, where the dataset sample n = 407,303.

Figure 4.12 shows sample ID vs. the number of observation and sequences, where the subphases number have an exponential graph, while the number of observations has linear time dependence. I'm interested in these samples as they represent the acquired data after Cassini's spacecraft was inserted into Saturn orbit.

Data collecting was carried out within various phases as well as sub-phases across the outer space mission. Every sub-phase also contains many sequences to which every sequence involves a specific figure or quantity of observations based on the judgment of the Cassini imaging team mission and other related teams such as engineering. The observation includes a series of samplings at which the size of the set counts on scientific events of the orbiter Cassini or outer space events related to the followed path.

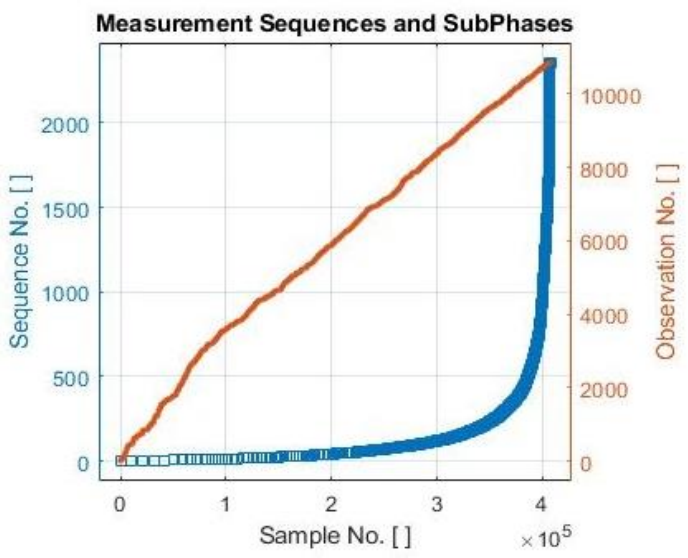


*Figure 4.12: Measurements of sequences and subphase vs. sample number*

In addition to the number of observations in the inspected time interval, the sequences were 2,355 and M = 10,850, respectively. Potential complex events of the data set are in the moment of change of observation sequences.

Within the conducted examination, I utilized a small proportional quantity but a representative group of independent variables ($k = 8$, check Table 4.4).

The supervised learning process for specifying the CKE technique's operating values for memory size r, the sensitivity of the outlier detection method ρ, and severity threshold $Th$ of the method I used first (1%) subset of the total samples ($n_{learn} = 407{,}3$). The remainder of (99%) of the data were utilized as test data to validate the CKE analysis model.

*Table 4.4: The analyzed list of the independent variable*

| Var. | Variable Name | Description |
|---|---|---|
| $v_1$ | Sampling interval [s] | Time int. between consecutive samplings |
| $v_2$ | Sampling durations [s] | Duration of samplings |
| $v_3$ | Detector temperature [°C] | Temp. of Charged Coupled Device (CCD) |
| $v_4$ | Filter temperature [°C] | Temperature of Wheels filters |
| $v_5$ | Diff. No. of packets/image | Abs(expected – received) packets/image |
| $v_6$ | Data Transmission Rate [kbit/s] | Rate of transmitted data by the orbiter |
| $v_7$ | Image compression ratio [bit/pixel] | Received images compression ratio |
| $v_8$ | Picture exposure time [s] | Exposure time of picture |

Sampling intervals and sampling durations are given in Figure 4.13 and Figure 4.14, respectively. Sampling interval and sampling duration will satisfy the time domain necessity. Where the horizontal axis shows the sample number, and the vertical axis represents the time.

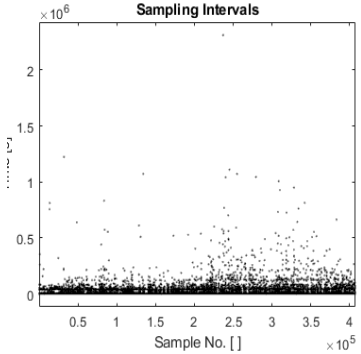


*Figure 4.13: Sampling intervals at the Cassini orbiter*

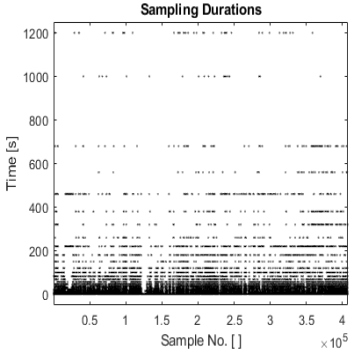


*Figure 4.14: Sampling durations at the Cassini orbiter*

The sampling time shows the time interval among consecutive samples, while the sampling duration means the difference in time among two successive samples. Under the varied ranges of the temperature for the detector and the filter, it is possible to detect special events by detecting the set of discrete points in the represented data. It can be noticed that the indispensable sampling time is around 15 s or more. The model provides the ability to track several observations. In practice, the expected packets and received packets for each image are different because of the image sampling quality. Radio signals shall need between (68 – 84) minutes to make their journey across the space separating Cassini and the ground station on the Earth. It is noticed that the data transmission rate is varied based on the active phase of the mission as the transfer rate ranges between 5 kilobits per second at the minimum rate, while the maximum rate is (365) kilobits per second. The content influences picture exposure time and compression of sampled data at the Cassini orbiter (see Figure 4.15 and Figure 4.16).

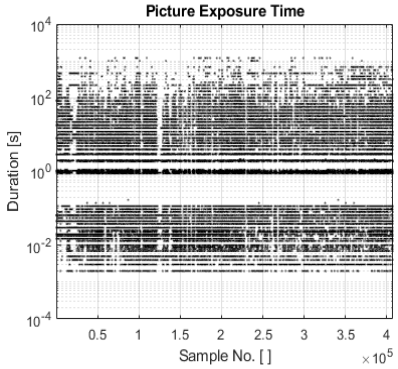


*Figure 4.15: Picture exposure time at Cassini orbiter*

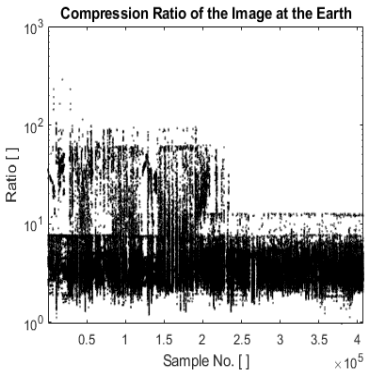


*Figure 4.16: Compression ratio of the received images*

The dependence of the CKE method hit ratio on parameters $r$ and $\rho$ are given in Figure 4.17, while the detected special events via the proposed method are presented in Figure 4.18. Through the analysis, it is also

expected that there will be several picture exposure times ranging from (0-1200) seconds. Short exposures are used needed to reduce smears throughout close Cassini flybys.

The compression ratio extent between 2 and 3 is determined by the data actual entropy. Moreover, based on the running activity, it incorporates different ratios ranging from (2:1) to the ratio of (10:1). The provided AI model qualified us to detect and confirm a considerable number of complex events based on these variations of ratios. For the learning time series of the C-H project, it was found that the optimum memory size is $r^* = 1 \cdot r_0 = 2$, and the optimum sensitivity of the outlier detector is $\rho^* = 0.7 \cdot \rho_0 = 0.7 \cdot (\max(V) - \min(V))/\sigma[n, r]$. The optimum severity threshold $Th$ of the CKE method was determined based on the requirement to find the nearest number of special events to the number of sequences ($M = 10{,}850$) executed during the project.

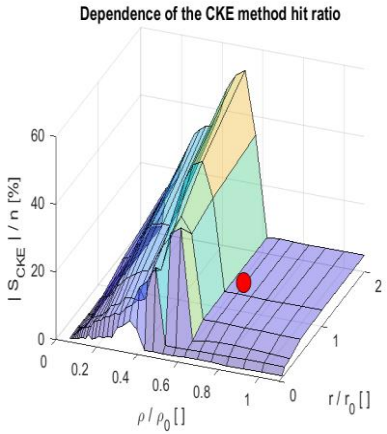


*Figure 4.17: Dependence of the CKE hit ratio on $r$ and $\rho$*

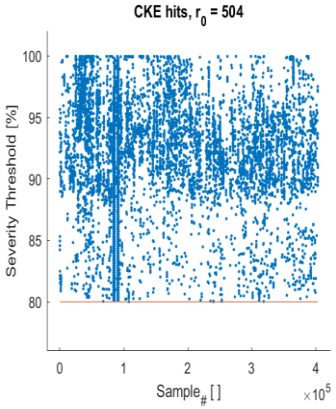


*Figure 4.18: Detected special events by CKE method*

The memory size and severity threshold were found to be( $r^* = 504$ and $Th = 87.6\ \%$), respectively. Resulting in $|\ S_{CKE}\ | = 10{,}134$ number of

special events detected in the testing data inducing hit ratio of the special events to be $| S_{CKE} | / n = 2.4\ \%$.

## 4.5. Meta-Learning LSTM Model

Thesis 5 (J3) (Section 5.5) This thesis speculates an essential topic related to the classification of remotely sensed images, in which the process of feature coding and extraction are decisive procedures. I put forward a straightforward, efficient classifier based on feature extraction by analyzing the cell of tensors via layered Map-Reduce framework beside meta-learning LSTM followed by a SoftMax classifier. Experiment results show that the provided model attains a classification accuracy of (96.7%), which makes the provided model quite valid for diverse image databases with varying sizes. The essential considered factors of quantitative image analysis are processing and analysis. Among the challenges that will face any researcher are software and hardware limitations. During our dataset processing and inspection, I encountered these kinds of restrictions. However, I overcame them; as I'm dealing with the Bd, I cannot depend on the regular computer hardware. I used Graphical Processing Units (GPUs); they are a hardware appliance that is most effective for parallel and rapid processing.

GPUs provide DL with the ability to perform separated computations from the central processor (that is, serial tasks dedicated) and adequately fulfilling complex computations. In our research, I will use the GPU to be eligible to process a big data set of Saturn images that contain more than 400,000 images. To initiate the process of training, I indiscriminately sampled 50K images from the adopted dataset volumes. Every image is correlated to one or more categories, arranged in 6 observable denominations, as represented in Figure 4.19. From left to right: Saturn, Rings, Titan, Icy Satellites, Small Satellites (rocks), Sky.

The label selection method is based on each image content, as it is delineated to an interpretative word just as Cassini teams adopted. Below is a screenshot of the classification process adopted above, which shows the image with its target with correct classification accuracy, even with not existing before images. The proposed algorithm is capable of processing at

high speed. Also, the effectiveness of the adopted approach to transfer learning is noticed. Table 4.5 shows the adopted classes and the number of images utilized in the primary processes of training and testing.

*Figure 4.19: Cassini mission images sampling presenting the adopted content labeling classes*

The framework of meta-learning could be executed to any technique which is trained via the scheme of meta-learning. The purpose is imparting an inclusive approach that can readily be ordered to achieve the best or the desired performance with any required task. The analyses of the cell of tensors are straightforward embedding via the layer of map-reducer discriminator (D), which is acting as a pooling layer that reduces the mapped features. At the same time, the meta-learning LSTM will deconstruct each generated tensor into four cell tensors: (input tensor cell), (forget tensor cell), (cell state tensor cell), and (output tensor cell).

The model complexity is determined by the revelation of Levin complexity definition given by the following formula [9]:

$C_L(P) = \min_P \{cL(p)$: if program p solves P and then ceases during time $t_P\}$ (4-2)

where $C_L(P) = 1(p) + log(t(p))$. (4-3)

The problem that needs to be solved is represented by P, while l(p) is the program p length, and t(p) represents the time that is consumed by p to solve P. Transferring knowledge acquired from a single task with the abundance of labeled data to some other tasks with slight labeled data, the level of

progression of performing its mission relies on how pertinent is the former task of big-scale image recognition to the current task [10].

*Table 4.5: Number of used images for the initiated process of training and testing*

| **Class** | Saturn | Rings | Titan | Icy Satellites | Small Satellites (rocks) | Sky |
|---|---|---|---|---|---|---|
| Training Images | 9865 | 8369 | 9568 | 7460 | 7879 | 6859 |
| Testing images | 1983 | 1870 | 2120 | 1987 | 2060 | 1580 |

In the situation of meta-learning LSTMs, with the epilogue of each task, the experience is gained and kept in the memory of the LSTM cell. This model is set to weigh with other commonly known image classification methods to confirm the proposed model efficiency. The conducted experiment results are demonstrated in Table 4.6, where it is evident that (Support Vector Machine) SVM and random forest models notably have less accuracy than the other models. The justification of this lies in the emphatic feature extraction performed by the provided model. Also, the proposed model adopts DL within its optimal optimization and parameter initialization.

*Table 4.6: A comparison among prior relevant work*

| **Model Name** | **Accuracy** |
|---|---|
| SVM [11] | 52.6 % |
| Random Forest [11] | 72.3 % |
| Fuzzy Clustering [12] | 88.78 % |
| Optimized Fuzzy system [13] | 93.07% and 95.25% |
| Sweep Image Transformation Technique [14] | 93.34% |
| Gray Level Co-occurrence Matrices [15] | 90% |
| **The Proposed Model** | **96.7 %** |

Figure 4.20 shows the diagram of training accuracy versus validation accuracy over the number of epochs.

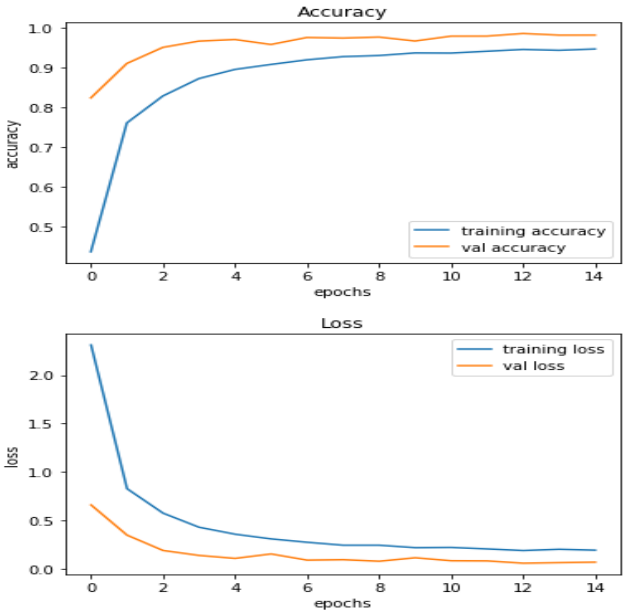


*Figure 4.20: Plot of model accuracy and loss on train, validation*

By measuring the model's accuracy, I can determine how effective it is performing. To get to desirable accuracy, I choose to decrease loss across the network, but not only on the validation range. When it comes to addressing classification issues, the confusion matrix is a highly common metric to employ [16]. A confusion matrix is a graphical representation of a classifier's classification performance concerning a set of test data [17]. The confusion matrix for calculating the overall classifier accuracy is shown in Figure 4.21, which is evaluated using the 50 thousand images dataset. The produced errors may be evaluated by a given model with a confusion matrix or its confusion degree.

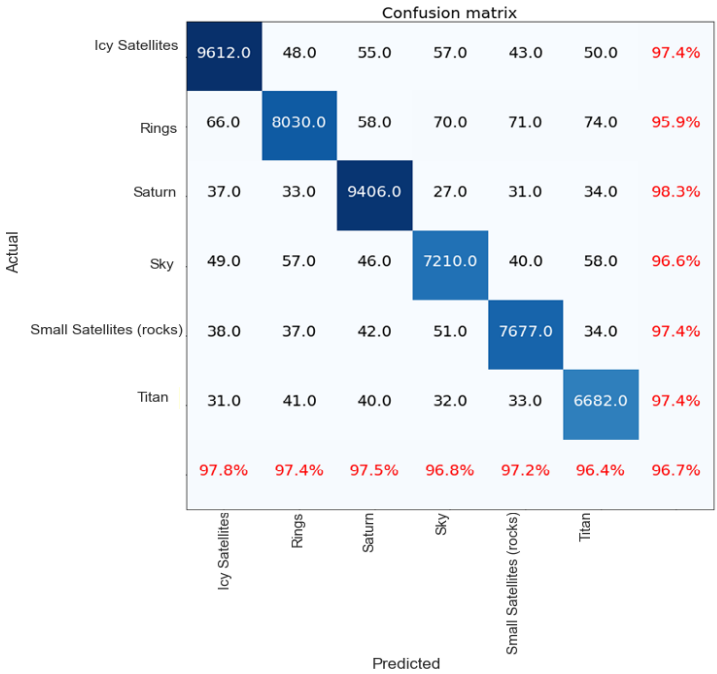


*Figure 4.21: Plot representation of the model confusion matrix*

If the model gets confused, the confusion matrix can record the confusion. The acquired percentage is (96.7%). The confusion matrix is a particular table structure that actually describes the algorithm's efficiency. Each column depicts the instances in a real class, while Every row in the matrix describes the instances in a class expected (or vice versa).

## 4.6. Deep Learning Model of Detecting Trajectory Modifications

Thesis 6 (J4, J5, C4, C5) (Section 5.6) This thesis provides a sophisticated in-depth learning approach for detecting Cassini spacecraft trajectory

modifications in post-processing mode. The model utilizes the ability of Recurrent neural Networks (RNNs) for drawing out useful data and learning the time series inner data pattern, along with the forcefulness of SNN layers for distinguishing dependencies among the long-short term. The research study exploited the statistical rates (MCC, F1) to evaluate the models. The preliminary analysis showed that exploiting the RNN layer provides a notable boost in rising the detection process performance. The proposed model achieved a number of 232 trajectory modification detections with (99.98 %) accuracy among the last 13.5 years of the Cassini spacecraft life.

The resulting learning time and the detection accuracy of these networks are given in Table 4.7. DL models specifically learn nuanced, task-adaptive, and high-level data functions. The results obtained show that a DL approach applied is an excellent approach with high accuracy.

As input data travels through the classifier, the information produced in each layer may be thought of as representations of the inputs in a particular dimensional space [18]. There were executed training and evaluation of the test data sets of the trajectory in the last 13.5 years of the Cassini project with three types of RNNs with six different hyperparameter settings for each of them, which makes eighteen neural networks in total. Six of these RNN are LSTM, and six are BiLSTM networks, while the last six are GRU.

Visualization of the loss during the learning process is represented in Figure 4.22. The results obtained show that a deep learning approach applied is an excellent approach with high accuracy. The changing input parameters (mini-batch size, number of hidden units, number of classes on layer three, and layer four). The less is the number of hidden units, the higher is the loss. For ten hidden units, a double learning loss is obtained. Similar behavior has it the RNN networks, it can be noticed that RNN12 becomes over learned for 100 hidden units on the L2 layer. Boosting data from training contributes to improve model accuracy [19]. Learning is done in streams of batches, which suggests that a number of training cases are sent across the network, and the results, known as forecasts, are gathered. [20] The batch size determines how many model parameters must be fit, and it determines how many predictions should be produced. [21] Using a bigger

batch size is more effective when trying to achieve greater accuracy. [22] It is emphasized that greater learning rates need bigger batch sizes.

*Table 4.7: Detection accuracy of the trajectory modification of different RNNs*

| System | RNN Type | Mini Batch Size [ ] | Hidden units# on L2 | Classes # on L3 | Classes # on L4 | Learni-ng time [s] | Detection accuracy [%] |
|---|---|---|---|---|---|---|---|
| $RNN_1$ | LSTM | 2500 | 10 | 100 | 2 | 799.5 | 97.42 |
| $RNN_2$ | LSTM | 2500 | 100 | 100 | 2 | 1352.5 | 97.13 |
| $RNN_3$ | LSTM | 5000 | 10 | 100 | 2 | 965.1 | 98.58 |
| **$RNN_4$** | **LSTM** | **5000** | **100** | **100** | **2** | **800.2** | **99.19** |
| $RNN_5$ | LSTM | 10000 | 10 | 100 | 2 | 1469.9 | 98.63 |
| $RNN_6$ | LSTM | 10000 | 100 | 100 | 2 | 992.4 | 98.97 |
| $RNN_7$ | BiLSTM | 2500 | 10 | 100 | 2 | 802.9 | 99.65 |
| $RNN_8$ | BiLSTM | 2500 | 100 | 100 | 2 | 1489.6 | 96.43 |
| **$RNN_9$** | **BiLSTM** | **5000** | **10** | **100** | **2** | **1014.0** | **99.75** |
| $RNN_{10}$ | BiLSTM | 5000 | 100 | 100 | 2 | 817.5 | 98.09 |
| $RNN_{11}$ | BiLSTM | 10000 | 10 | 100 | 2 | 1497.4 | 99.68 |
| $RNN_{12}$ | BiLSTM | 10000 | 100 | 100 | 2 | 1175.8 | 99.90 |
| $RNN_{13}$ | GRU | 2500 | 10 | 100 | 2 | 986.5 | 99.92 |
| $RNN_{14}$ | GRU | 2500 | 100 | 100 | 2 | 818.9 | 99.98 |
| $RNN_{15}$ | GRU | 5000 | 10 | 100 | 2 | 1015.8 | 99.63 |
| **$RNN_{16}$** | **GRU** | **5000** | **100** | **100** | **2** | **817.8** | **99.98** |
| $RNN_{17}$ | GRU | 10000 | 10 | 100 | 2 | 1550.3 | 97.60 |
| $RNN_{18}$ | GRU | 10000 | 100 | 100 | 2 | 1036.7 | 99.95 |

The loss is measured and recorded for each batch, which shows the total mistake. It is evident if accuracy improves, the loss decreases. The ideal learning rate is bound to be subjected to the loss of the leaned behavior of the data, which in its role is relying on together the used dataset and the architecture of the model. [23] As when dealing with data streams, batch learning is suitable. The value of the loss denotes how deficiently or efficiently the model is conducting after each iteration of learning. The value of the loss denotes how deficiently or efficiently the model is conducting after each iteration of learning.

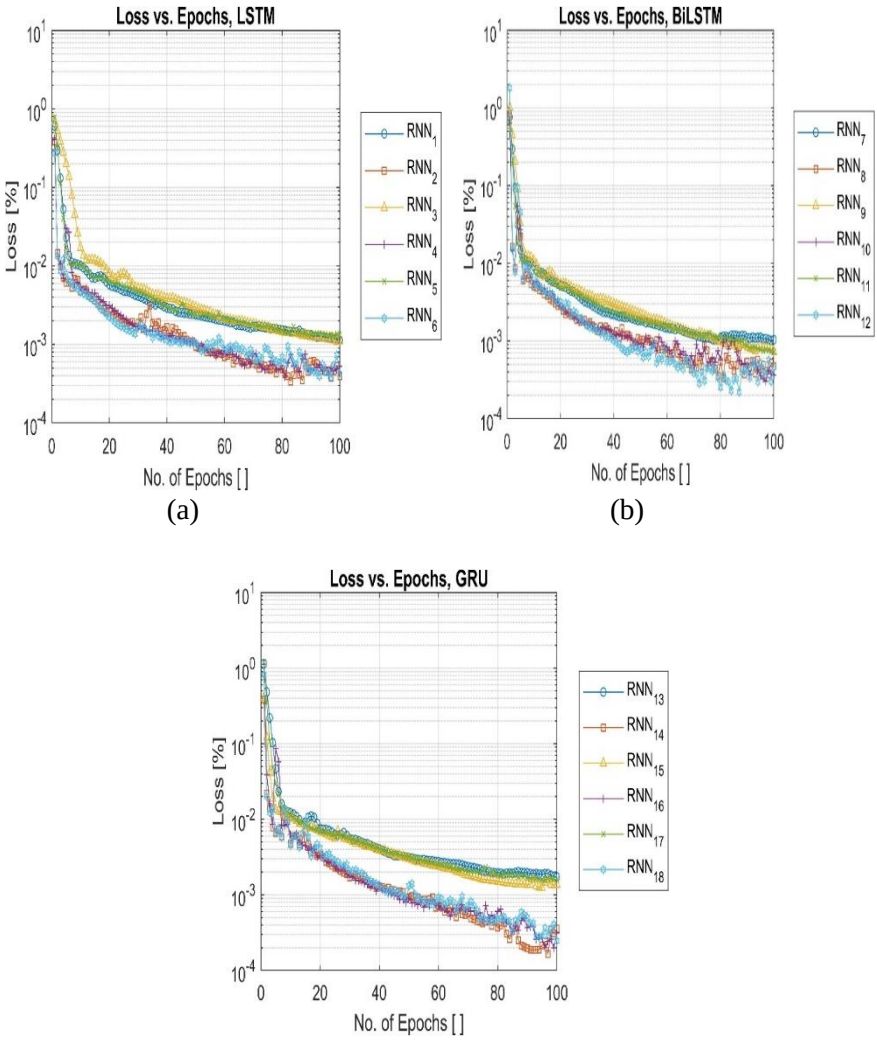


*Figure 4.22: Dependence of the minibatch learning loss on the number of epochs of the RNNs. (a): LSTM; (b): BiLSTM (c): GRU*

It's clear that the amount of hidden units on L2 has an effect on learning loss. The loss increases as the amount of hidden units decreases. The learning deficit is doubled when there are ten hidden units. Analogous feature is observed on the LSTM, BiLSTM and GRU networks, but the last two seems to become over-learned after utilizing one hundred hidden units on the layer of L2. I have trained the framework until it has completed 100 epochs and displayed the training loss values vs. the epochs number.

As the amount of epochs rises, the loss of training is reduced. As compared to other LSTM networks, RNN4 LSTM has the lowest loss. RNN9 is offering the highest outcomes in terms of losses among the other BiLSTM networks, and GRU16 is offering the highest outcomes in terms of losses among the other GRU networks. The ideal learning rate is unavoidably revealed to the data's lack of learned behaviour which is dependent on both the dataset and the model's design. The loss value shows how successfully or incompetently the model works during each learning iteration. It is evident in the Figure 4.23a, that RNN4 LSTM and RNN16 GRU are offering the ideal accuracy when set side by side to other RNN network models.

It depicts the relationship between accuracy and learning period with each of the RNNs studied. Most of the GRU network need less time to learn than BiLSTM and LSTM networks. The accuracy metric is utilized to quantify the algorithm functioning in an explicable method. Generally, the model accuracy is specified after calculating the used model parameters with a percentage style. It is considered as the model accuracy measure that represents the model prediction contrasted to the correct data. The accuracy measure is used to calculate the model in an explainable way. The accuracy of the model is usually is based on the values that are calculated by using the model parameters are defined. It is known as the algorithm accuracy metric, reflecting the algorithm forecast in reference to the right data. All LSTM networks need lower learning time than BiLSTM networks. The dependency of the $F_1$ factor on the $MCC$ coefficient is given in Figure 4.23b. Having different calculation formula of performance metrics, I used multiple metrics to evaluate relation between detection accuracy, $MCC$ and $F_1$ score of the analysed data series.

Model precision implies the number of classifications that a trend accurately forecasts, divided by the total number of predictions produced by a model correctly predicts. It is a way to measure mode results. The model accuracy is identified after calculating the used model parameters with a percentage style; this is clear in the figure below.

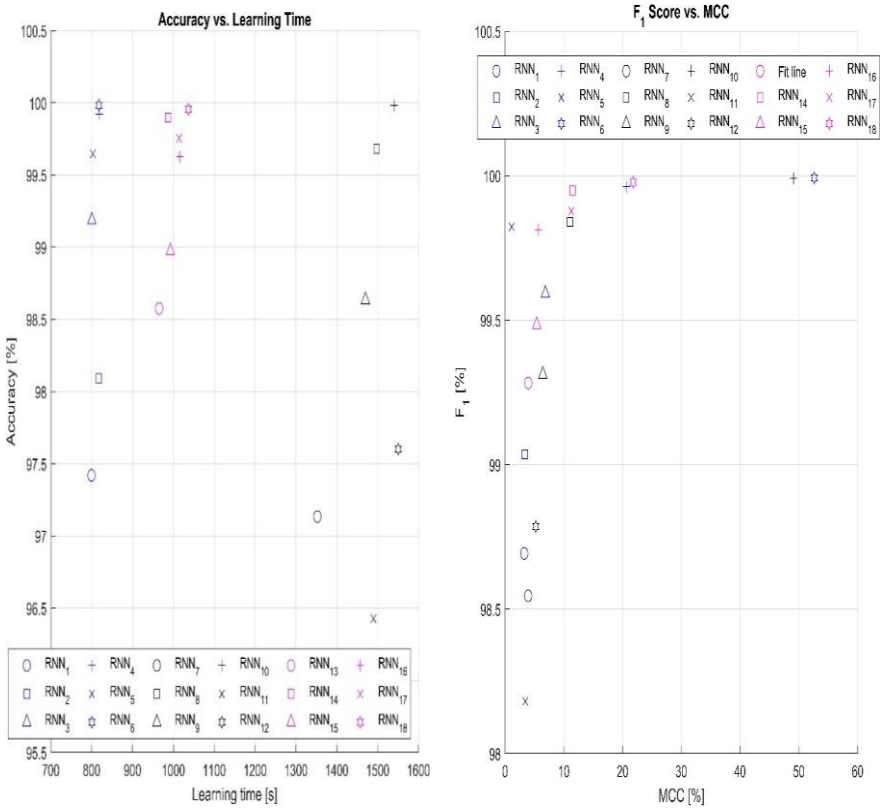


*Figure 4.23: Scatter plot of the learning time and accuracy of the RNNs (a); Dependency of the $F_1$ score on the Matthew correlation coefficient $MCC$ (b)*

As an illustration, RNN4 is hitting an accuracy that exceeds the value of 99.9 % with a minimum learning time. Mathew correlation coefficient $MCC$ and $F_1$ score of the resulting neural networks is given in Table 4.8. The following is a list of RNNs in tapering order for the MCC metric: RNN6, RNN10, RNN18, RNN4, RNN17, RNN8, RNN3, RNN9, RNN16, RNN16, RNN15, RNN12, RNN13, RNN7, RNN11, RNN2, RNN1, RNN5.

Through observing this outcome, it can be concluded that the optimal RNN to recognize the modification among Cassini trajectory is GRU accompanied by 100 hidden units and a MiniBatch of 5000 and is able to model trajectory change behaviour with 99.98% precision, with very little time fewer than 13.7 minutes on a desktop machine with 12 central processing cores, and 64 GB of RAM.

*Table 4.8: Mathews correlation coefficient MCC and F_1 score of the analyzed RNNs*

| **RNN** | **1** | **2** | **3** | **4** | **5** | **6** |
|---|---|---|---|---|---|---|
| Type | **LSTM** | | | | | |
| $MCC$ [%] | 3.27 | 3.36 | 6.85 | 20.69 | 1.12 | 52.63 |
| $F_1$ [%] | 98.69 | 99.04 | 99.59 | 99.96 | 99.82 | 99.99 |
| | | | | | | |
| **RNN** | **7** | **8** | **9** | **10** | **11** | **12** |
| Type | **BiLSTM** | | | | | |
| $MCC$ [%] | 3.92 | 11.06 | 6.42 | 49.09 | 3.47 | 5.22 |
| $F_1$ [%] | 98.55 | 99.84 | 99.31 | 99.99 | 98.18 | 98.79 |
| | | | | | | |
| **RNN** | **13** | **14** | **15** | **16** | **17** | **18** |
| Type | **GRU** | | | | | |
| $MCC$ [%] | 3.94 | 11.46 | 5.38 | 5.69 | 11.28 | 21.80 |
| $F_1$ [%] | 99.28 | 99.95 | 99.48 | 99.81 | 99.88 | 99.99 |

# List of Own Publications

## List of Journals Papers [J]

**J1** **Ashraf ALDabbas,** Zoltán Gál: Learning and Reasoning with structured prediction Based on Revealing Event Complexity, International Journal of Advanced Science and Technology, pp 13816 - 13828 Vol. 29 No. 3, 2020. ISSN:22076360,20054238.http://sersc.org/journals/index.php/IJAST/article/view/31723. (Q4: **IF 0.41**)

**J2** **Ashraf ALDabbas,** Zoltan Gal: On the constructive knowledge-based event intelligent identification mechanism, Journal of Theoretical and Applied Information Technology. Vol 98 December 2020, Issue 24, 2020. ISSN: 18173195,19928645.http://www.jatit.org/volumes/Vol98No24/18Vol98No24.pdf. (Q3: IF 0.628)

**J3** **Ashraf ALDabbas**, Zoltan Gal: Cassini-Huygens Mission Images Classification Framework by Deep Learning Advanced Approach, *International Journal of Electrical and Computer Engineering.* Vol 11, No 3,2020.ISSN2088-8708. DOI: http://doi.org/10.11591/ijece.v11i3.pp2457-2466**. (Q2: IF 2.3)**

**J4** **Ashraf ALDabbas**, Zoltán Gál: Detection Model of Cassini-Huygens Orbit Modifications Based on Recurrent Neural Networks, *Journal of Information and Communication Technology.* (Submitted) in the review process. ISSN: 21803862, 1675414X **(Q2: IF 1.8)**

**J5 Aldabbas, Ashraf**, Zoltan Gal, Khawaja Moyeezullah Ghori, Muhammad Imran, and Muhammad Shoaib. "Deep Learning-Based Approach for Detecting Trajectory Modifications of Cassini-Huygens Spacecraft." *IEEE Access* 9 (2021). pp. 39111-39125. ISSN: 21693536, DOI: 10.1109/access.2021.3064753 . **(Q1: IF 3.745)**

**J6 ALDabbas Ashraf**, Zoltan Gal, and Buchman Attila. "Neural Network Estimation of Tourism Climatic Index (TCI) Based on Temperature-Humidity Index (THI)-Jordan Region Using Sensed Datasets." *Carpathian Journal of Electronic and Computer Engineering* 11, no. 2. pp. 50-55, 2018.

**J7** Jihad Fraij, **Ashraf ALDabbas**, Nemer Aburumman. Blockchain as an E-Voting Tool. *International Journal of Advanced Research* 8(12):858-866, 2020. ISSN 2320-5407, DOI: 10.21474/IJAR01/12225.

## List of Conference Proceedings [C]

**C1 ALDabbas Ashraf,** and Zoltán Gál. "Getting facts about interplanetary mission of Cassini-Huygens spacecraft." In *10th Hungarian GIS Conference and Exhibition, Debrecen, Hungary*. 2019.

**C2 ALDabbas Ashraf**, and Zoltán Gál. "Complex Event Processing Based Analysis of Cassini–Huygens Interplanetary Dataset." In *International Conference on Information, Communication and Computing Technology*, pp. 51-66. Springer, Cham, 2019. DOI: https://doi.org/10.1007/978-3-030-38501-9_5.

**C3 ALDabbas Ashraf,** and Zoltan Gal. "On the Complex Event Identification Based on Cognitive

Classification Process." In *2019 10th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, pp. 29-34. IEEE, 2019.

**C4** **ALDabbas Ashraf**, and Zoltán Gál. "Change Detection of the Cassini Orbit Based on Data Dissimilarity" In 11th Hungarian GIS Conference and Exhibition, Debrecen, Hungary. 2020.

**C5** **ALDabbas Ashraf**, and Zoltán Gál. "Deep Learning Based Approach for Detecting Cassini-Huygens Spacecraft Trajectory Modifications" In The 1st Conference on Information Technology and Data Science, Debrecen, Hungary. 2020.

**C6** Gal, Zoltan, Mohamed Amine Korteby, and **Ashraf ALDabbas**. "Impact of the delay tolerance in wireless sensor networks: Buffer occupancy and energy consumption aspects." In 2018 IEEE International Conference on Future IoT Technologies (Future IoT), pp. 1-8. IEEE, 2018.

**C7** **Ashraf ALDabbas** – Zoltán Gál – Mohamed Amine Korteby: 3D GIS - A Major Step Analysis to Evaluate Convenient Healthy Residential District Based on Environmental Sensory Data Sets, *9th Hungarian GIS Conference and Exhibition, Debrecen, Hungary.* May 24-25, 2018.

**C8** Mohamed Amine Korteby – Zoltán Gál – **Ashraf Dabbas**: Impact of the Geographic Map-Based Movement on the Communication Quality of Sensor Networks, *9th Hungarian GIS Conference and Exhibition*. Debrecen, Hungary. May 24-25, 2018,

**C9** Gal, Zoltan, Mohamed Amine Korteby, and **Ashraf ALDabbas**. "Estimation of the Wireless Sensor Network Performance using Fractal Behavior of the

Generated Cognitive Harmonic Waves." In *2018 9th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, pp. 57-62. IEEE, 2018.

# Bibliography

[1] Karpatne, Anuj, Imme Ebert-Uphoff, Sai Ravela, Hassan Ali Babaie, and Vipin Kumar. "ML for the geosciences: Challenges and opportunities." IEEE Transactions on Knowledge and Data Engineering 31, no. 8, pp. 1544-1554, 2018.

[2] Rudin, Cynthia. "Stop explaining black box ML models for high stakes decisions and use interpretable models instead." Nature Machine Intelligence 1, no. 5, pp.206-215, 2019.

[3] C. Luo, D. Yang, J. Huang, and Y.-D. Deng. "LSTM-Based Temperature Prediction for Hot-Axles of Locomotives". In: Proc. of International Conference on Information Technology and Applications (ITA). Beijing, China, June 2017, pages 1–6

[4] Q. Zhang, H. Wang, J. Dong, et al. "Prediction of Sea Surface Temperature Using Long Short-Term Memory". In: Journal of IEEE Geoscience and Remote Sensing Letters 14.10 (Oct. 2017), pages 1745–1749.

[5] Marsland, S. ML (CRC Press, Taylor & Francis Inc., Boca Raton, FL.) 2014.

[6] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Delving deep into rectifiers: Surpassing human-level performance on imagenet

classification." In Proceedings of the IEEE international conference on computer vision, pp. 1026-1034. 2015.

[7] Liu, Shuai-shi, and Yan-tao Tian. "Facial expression recognition method based on gabor wavelet features and fractional power polynomial kernel PCA." In International Symposium on Neural Networks, pp. 144-151. Springer, Berlin, Heidelberg, 2010.

[8] Bojarski, Mariusz, Davide Del Testa, Daniel Dworakowski, Bernhard Firner, Beat Flepp, Prasoon Goyal, Lawrence D. Jackel et al. "End to end learning for self-driving cars." arXiv preprint arXiv:1604.07316 (2016).

[9] M.Li and P. Vitányi,"An introduction to Kolmogorov complexity and its applications,"New York: Springer, vol. 3, 2008.

[10] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in Advances in neural information processing systems, pp. 3320–3328, 2014.

[11] L. Zhu and P. Spachos, "Towards Image Classification with ML Methodologies for Smartphones," ML and Knowledge Extraction, vol. 1, no. 4, pp. 1039–1057, 2019.

[12] F. Yan, W. Mei, and Z. Chunqin, "SAR image target recognition based on Hu invariant moments and SVM," in 2009 Fifth International Conference on Information Assurance and Security, vol. 1, pp. 585–588, 2009.

[13] M. Ramezanifard and B. S. Mousavi, "Digital

image classification by optimised fuzzy system," Indonesian Journal of Electrical Engineering and Computer Science, vol. 14, no. 3, pp. 1196–1202, 2019.

[14] S. Ibrahim, N. A. Zulkifli, N. Sabri, A. A. Shari, and M. R. M. Noordin, "Rice grain classification using multi-class support vector machine (SVM)," IAES International Journal of Artificial Intelligence, vol. 8, no. 3, 2019.

[15] Y. Sari, P. B. Prakoso, and A. R. Baskara, "Application of neural network method for road crack detection," TELKOMNIKA Telecommunication Computing Electronics and Control, vol. 18, no. 4, pp. 1962–1967, 2020.

[16] Kulkarni, Ajay, Deri Chong, and Feras A. Batarseh. "Foundations of data imbalance and solutions for a data democracy." In *Data Democracy*, Academic Press, pp. 83-106, 2020.

[17] Ting, Kai Ming. "Confusion matrix." *Encyclopedia of Machine Learning and Data Mining* 260, 2017.

[18] Hinton, Geoffrey E. "Learning multiple layers of representation." *Trends in cognitive sciences* 11. No. 10: pp. 428-434. 2007.

[19] Lei, Suhua, Huan Zhang, Ke Wang, and Zhendong Su. "How training data affect the accuracy and robustness of neural networks for image classification." 2018.

[20] Radiuk, Pavlo M. "Impact of training set batch size on the performance of convolutional neural networks for diverse datasets." *Information*

*Technology and Management Science* 20, no. 1, pp.20-24, 2017.

[21] Devarakonda, A., M. Naumov, and M. Garland. "AdaBatch: Adaptive batch sizes for training deep neural networks. arXiv 2017." *arXiv preprint arXiv:1712.02029*.

[22] Kandel, Ibrahem, and Mauro Castelli. "The effect of batch size on the generalizability of the convolutional neural networks on a histopathology dataset." *ICT express* 6, no. 4, pp.312-315, 2020.

[23] Read, Jesse, Albert Bifet, Bernhard Pfahringer, and Geoff Holmes. "Batch-incremental versus instance-incremental learning in dynamic and evolving data." In *International symposium on intelligent data analysis*, pp. 313-323. Springer, Berlin, Heidelberg, 2012.

UNIVERSITY of DEBRECEN

UNIVERSITY AND NATIONAL LIBRARY
UNIVERSITY OF DEBRECEN
H-4002 Egyetem tér 1, Debrecen
Phone: +3652/410-443, email: publikaciok@lib.unideb.hu

Registry number: DEENK/395/2021.PL
Subject: PhD Publication List

Candidate: Ashraf Khaled Abd Elkareem Aldabbas
Doctoral School: Doctoral School of Informatics
MTMT ID: 10077449

## List of publications related to the dissertation

Foreign language scientific articles in international journals (5)

1. **Aldabbas, A. K. A. E.**, Gál, Z.: Cassini-Huygens mission images classification framework by deep learning advanced approach.
   *Int. J. Electr. Comp. Eng. 11* (3), 2457-2466, 2021. ISSN: 2088-8708.
   DOI: http://dx.doi.org/10.11591/ijece.v11i3.pp2457-2466
2. **Aldabbas, A. K. A. E.**, Gál, Z., Ghori, K. M., Imran, M., Shoaib, M.: Deep Learning-Based Approach for Detecting Trajectory Modifications of Cassini-Huygens Spacecraft.
   *IEEE Access. 9*, 39111-39125, 2021. ISSN: 2169-3536.
   DOI: http://dx.doi.org/10.1109/ACCESS.2021.3064753
   IF: 3.367 (2020)
3. **Aldabbas, A. K. A. E.**, Gál, Z.: Learning and Reasoning with structured Prediction Based on Revealing Event Complexity.
   *Int. J. Adv. Sci. Technol. 29* (3), 13816-13828, 2020. ISSN: 2005-4238.
4. **Aldabbas, A. K. A. E.**, Gál, Z.: On the constructive knowledge-based event intelligent identification mechanism.
   *J. Theor. Appl. Inf. Technol. 98* (24), 4169-4180, 2020. ISSN: 1992-8645.
5. **Aldabbas, A. K. A. E.**, Gál, Z., Buchman, A.: Neural Network Estimation of Tourism Climatic Index (TCI) Based on Temperature-Humidity Index (THI)- Jordan Region Using Sensed Datasets.
   *Carpathian J. Electr. Comp. Eng. 11* (2), 50-55, 2018. ISSN: 1844-9689.
   DOI: http://dx.doi.org/10.2478/cjece-2018-0019

Foreign language conference proceedings (6)

6. **Aldabbas, A. K. A. E.**, Gál, Z.: Change Detection of the Cassini Orbit Based on Data Dissimilarity.
   In: Az elmélet és a gyakorlat találkozása a térinformatikában XI. : theory meets practice in GIS. Szerk.: Molnár Vanda Éva, Debreceni Egyetemi Kiadó, Debrecen, 23-29, 2020. ISBN: 9789633188866

**UNIVERSITY AND NATIONAL LIBRARY**
**UNIVERSITY OF DEBRECEN**
H-4002 Egyetem tér 1, Debrecen
Phone: +3652/410-443, email: publikaciok@lib.unideb.hu

7. **Aldabbas, A. K. A. E.**, Gál, Z.: Complex Event Processing Based Analysis of Cassini-Huygens Interplanetary Dataset.
In: Intelligent Computing Paradigm and Cutting-edge Technologies. Ed.: Lakhmi C. Jain, Sheng-Lung Peng, Basim Alhadidi, Souvik Pal, Springer, Cham, 51-66, 2020, (Learning and Analytics in Intelligent Systems, ISSN 2662-3447 ; 9) ISBN: 9783030385002

8. **Aldabbas, A. K. A. E.**, Gál, Z.: Deep Learning-Based Method for Detecting Cassini-Huygens Spacecraft Trajectory Modifications.
In: Proceedings of the 1st Conference on Information Technology and Data Science. Eds.: István Fazekas, András Hajdu, Tibor Tómács, CEUR, Debrecen, 19-31, 2020, (CEUR Workshop Proceedings, ISSN 1613-0073 ; 2874)

9. **Aldabbas, A. K. A. E.**, Gál, Z.: Getting facts about interplanetary mission of Cassini-Huygens spacecraft.
In: Az elmélet és a gyakorlat találkozása a térinformatikában X. = theory meets practice in GIS. Szerk.: Molnár Vanda Éva, Debrecen Egyetemi Kiadó, Debrecen, 27-34, 2019.

10. **Aldabbas, A. K. A. E.**, Gál, Z.: On the Complex Event Identification Based on Cognitive Classification Process.
In: 10th IEEE International Conference on Cognitive Infocommunications CogInfoCom 2019 : Proceedings. Ed.: Péter Baranyi, Anna Esposito, Nelson Mauro Maldonato, Carl Vogel, IEEE, Piscataway, 29-34, 2019. ISBN: 9781728147932

11. **Aldabbas, A. K. A. E.**, Gál, Z., Korteby, M. A.: 3D GIS - A Major Step Analysis to Evaluate Convenient Healthy Residential District Based on Environmental Sensory Data Sets.
In: Az elmélet és a gyakorlat találkozása a térinformatikában IX. = Theory meets practice in GIS. Szerk.: Molnár Vanda Éva, Debreceni Egyetemi Kiadó, Debrecen, 77-84, 2018. ISBN: 9789633187234

## List of other publications

Foreign language international book chapters (1)

12. Gál, Z., Korteby, M. A., **Aldabbas, A. K. A. E.**: Impact of the delay tolerance in wireless sensor networks: Buffer occupancy and energy consumption aspects.
In: Proceedings of 2018 IEEE International Conference on Future IoT Technologies (Future IoT 2018), IEEE, Washington, [1-8], 2018. ISBN: 9781538612088



Foreign language scientific articles in international journals (1)

13. Fraij, J. K. I., **Aldabbas, A. K. A. E.**, Aburumman, N. M. A.: Blockchain as an e-voting tool.
*Int. J. Adv. Res. 8* (12), 858-866, 2020. EISSN: 2320-5407.
DOI: http://dx.doi.org/10.21474/IJAR01/12225

UNIVERSITY of DEBRECEN

**UNIVERSITY AND NATIONAL LIBRARY**
**UNIVERSITY OF DEBRECEN**
H-4002 Egyetem tér 1, Debrecen
Phone: +3652/410-443, email: publikaciok@lib.unideb.hu

Foreign language conference proceedings (2)

14. Gál, Z., Korteby, M. A., **Aldabbas, A. K. A. E.**: Estimation of the Wireless Sensor Network Performance using Fractal Behavior of the Generated Cognitive Harmonic Waves. In: 9th IEEE International Conference on Cognitive Infocommunications: CogInfoCom 2018: Proceedings. Ed.: by Péter Baranyi, IEEE, Piscataway, 57-62, 2018. ISBN: 9781538670941
15. Korteby, M. A., Gál, Z., **Aldabbas, A. K. A. E.**: Impact of the Geographic Map Based Movement on the Communication Quality of Sensor Networks. In: Az elmélet és a gyakorlat találkozása a térinformatikában IX. = Theory meets practice in GIS. Szerk.: Molnár Vanda Éva, Debreceni Egyetemi Kiadó, Debrecen, 179-186, 2018. ISBN: 9789633187234

**Total IF of journals (all publications): 3,367**
**Total IF of journals (publications related to the dissertation): 3,367**

The Candidate's publication data submitted to the iDEa Tudóstér have been validated by DEENK on the basis of the Journal Citation Report (Impact Factor) database.

05 August, 2021

DEBRECENI EGYETEM * Egyetemi és Nemzeti Könyvtár *