Attention-based Long Short Term Memory Model for DNA Damage Prediction in Mammalian Cells

Mohammad A. Alsharaiah[1]*, Laith H. Baniata[2]*, Omar Adwan[3], Ahmad Adel Abu-Shareha[4], Mosleh Abu Alhaj[5]
Qasem Kharma[6], Abdelrahman Hussein[7], Orieb Abualghanam[8], Nabeel Alassaf[9] and Mohammad Baniata[10]
Al-Ahliyya Amman University; Amman, Jordan[1, 3, 4, 5, 6, 7, 8]
Gachon University; South Korea[2]
The University of Jordan[9]
Ubion, South Korea[10]

Abstract*—The understanding of DNA damage intensity – concentration-level is critical for biological and biomedical research, such as cellular homeostasis, tumor suppression, immunity, and gametogenesis. Therefore, recognizing and quantifying DNA damage intensity levels is a substantial issue, which requires further robust and effective approaches. DNA damage has several intensity levels. These levels of DNA damage in malignant cells and in other unhealthy cells are significant in the assessment of lesion stages located in normal cells. There is a need to get more insight from the available biological data to predict, explore and classify DNA damage intensity levels. Herein, the development process relied on the available biological dataset related to DNA damage signaling pathways, which plays a crucial role in DNA damage in the mammalian cell system. The biological dataset that was used in the proposed model consists of 15000 records intensity – concentration-level for a set of five proteins which regulate DNA damage. This research paper proposes an innovative deep learning model, which consists of an attention-based long short term-memory (AT-LSTM) model for DNA damage multi class predictions. The proposed model splits the prediction procedure into dual stages. For the first stage, we adopt the related feature sequences which are inserted as input to the LSTM neural network. In the next stage, the attention feature is applied efficiently to adopt the related feature sequences which are inserted as input to the softmax layer for prediction in the following frame. Our developed framework not only solves the long-term dependence problem of prediction effectively, but also enhances the interpretability of the prediction methods that was established on the neural network. We conducted a novel proposed model on big and complex biological datasets to perform prediction and multi classification tasks. Indeed, the (AT-LSTM) model has the ability to predict and classify the DNA damage in several classes: No-Damage, Low-damage, Medium-damage, High-damage, and Excess-damage. The experimental results show that our framework for DNA damage intensity level can be considered as state of the art for the biological DNA damage prediction domain.

Keywords*—Mammalian cell; deep learning techniques; attention; LSTM; classification; DNA damage

I. Introduction

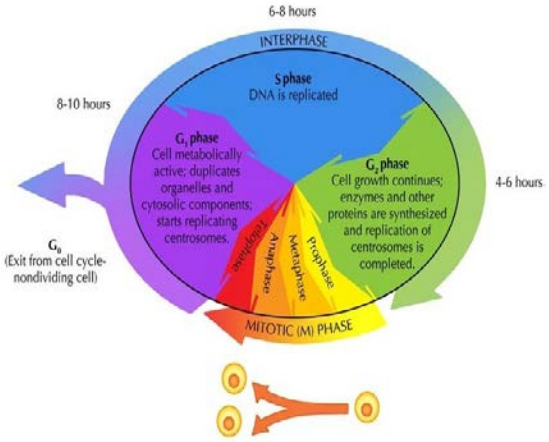
Mammalian cells have a complicated organism system. Specifically, each cell has a sequence of response procedures through a parent cell which is split into binary offspring cells; this is termed the cell-sequence-cycle with a total time of 24 hours. It consists of five phases, as shown in Fig. 1(a), Gap1 (G1) 8-10 hours, DNA synthesis (S) 6-8 hours, Gap2 (G2) 4-6 hours, Mitosis (M) around 4 hours and Quiescence (G0) silent mode. Furthermore, the mammalian cell has substantial impact on living cell dynamics, involving cell proliferation with differentiation [1]. However, mammalian cells usually stay in the early state or resting state, either Quiescence the G0 phase or initial G1phase, but the cell cycle developments to S phase further than the check point when actual growing influences motivate a cell necessarily. After DNA duplication through the S phase, the cell cycle developments complete the G2 phase to the final phase called the M phase. At the end of the cell cycle, specifically through M phase, the cell is necessarily separated into two new cells, called daughter cells. It signifies the complete progression process in the cell cycle as illustrated in Figure 1(a), (b). Also, this progression process is controlled by several complex networks. These networks enclose several biochemical species such as genes and proteins [2].

A mammal cell is commonly damaged and harmed by different resources like ultraviolet (UV)-irradiation, also ionization-radiation (IR), or other toxic chemical elements that are able to influence and cause breaks inside double-stranded DNA. This leads to DNA damage, and simulates an exceptional signal in the cell. Precisely, this DNA damage signal fires a DNA damage signaling pathway. The signaling pathway cooperates with the cell cycle controlling system to tentatively stop the cell cycle evolution in order to repair damaged DNA. Naturally, DNA damage is organized through a sub-network with five components, as shown in Fig. 2, and they cooperate over these steps (1) double significant elements at the launch are activated such as Ataxia telangiectasia mutated (ATM) and Rad3-related (ATR) protein kinases are activated via DNA damage, (2) ATM and ATR prompt p53 and checkpoint kinase 1 (Chk1), (3) initiated p53 stimulates the synthesis of p21, (4) then p21 prompting cell cycle halt. Accordingly, the signaling pathway for DNA damage takes straight action on the cell cycle arrangement mechanism, supporting cellular homeostasis and genetic constancy. Besides, any cell that has significant DNA damage might prompt apoptosis and perform planned cell death [3], [4].

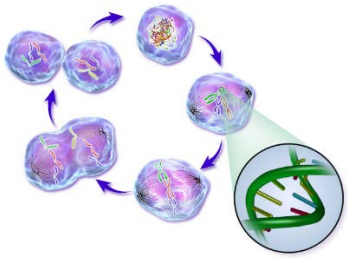
The influence of DNA damage in mammalian cells is one great cause of human illnesses, and as such has gained much interest in research since the mid-1990s. DNA damage and oxidative stress are identifying factors for the source,

*Corresponding Authors.

development, aetiology and progression of numerous different types of human disorders and diseases, such as cancer. Therefore, an abundance of present-day investigation in the DNA damage domain is dedicated towards sympathetic mechanisms and natural allegations of harmful DNA. This harmful DNA can go through alterations, such as mutations on its genetics; these types of mutations ultimately prime the expansion of tumors. DNA damage is also concerned in the growth of further prevalent human illnesses ranging from neurodegenerative disorders, such as Alzheimer's illness, to chronic obstructive pulmonary disease (COPD). Further, they have also been linked to diverse illnesses, such as pulmonary illnesses, brain injury and other chronic inflammation related to disorders [5].



(a) The Main Phases inside Mammalian Cell Cycle System.



(b) Illustration for Mammalian Cell Progression and Division Process

Fig. 1. Cell Cycle System Stages and Progression for Mammal Cell.

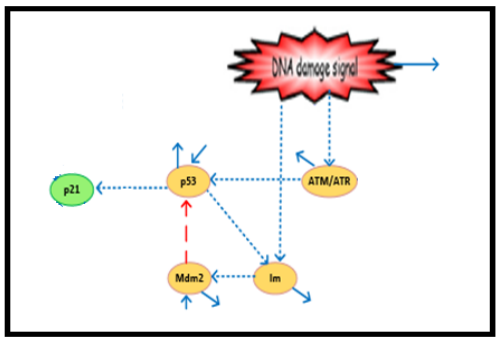


Fig. 2. The Main Elements for DNA Damage Signaling Pathway.

Essentially, biological discoveries have shown that mammalian cells are able to approximate the intensity-concentrations of DNA damage and choose a suitable cell fate, like applying DNA reparation; otherwise, cell cycle arrest, or apoptosis death. Nevertheless, it is uncertain the manner in which a cell decides the suitable cell destiny. A confirmation of the affiliation among the intensity (proteins concentration) for DNA damage and the energetic behavior of the biochemical elements implicated in cell cycle controlling techniques and the signaling pathway for DNA damage is crucial for clarifying the techniques of cell destiny purposes.

DNA damage has a number of intensity-concentration levels; these levels of DNA damage in malignant cells and in further diseased cells are significant in the assessment of the lesion stages that appear in normal cells. A wealth of laboratory research has been concerned with distinguishing and comprehending DNA damage levels and DNA repair capacity, as well as the techniques employed through mutually abnormal and normal cells. In addition, since certain significant diseases, such as cancer, are the essential reason of premature mortality over the globe, there is a predominance and rapid assertiveness of research on illustrative DNA damage in cancer cells [6]. Mainly, as mentioned before, clarifying the DNA damage level will be helpful for treatments and research purposes, even if there is a scarcity in the available data. Consequently, this article aims to present a novel artificial deep learning model to classify and predict DNA damage levels based on available DNA damage intensity-concentration levels from a bench mark model. The bench mark model delivers a novel dataset for DNA signaling network and classifies the DNA damage into several levels. We rely on this dataset to train and test the novel proposed model to predict the weather of the DNA damage and classify them in several classes.

The research introduces and validates the novel model to predict and classify DNA damage levels. To achieve the goals, outcomes have been investigated with DNA damage datasets. The research objectives and contributions are represented as follows:

1) This research aims to propose a novel deep learning model by employing an Attention – Long Short Term Memory.

2) Experiments are to be applied on DNA damage datasets.

3) The ATT-LSTM deep learning classifier for DNA damage is to be employed, and the efficiency of the ATT-LSTM deep learning classifier is to be determined.

4) The developed framework not only solves the long-term dependence problem of prediction effectively, but also enhances the interpretability of the prediction methods established on the neural network.

The paper is structured as follows. Section II offers a literature review. Section III encloses the utilized method and model architecture and implementation. Section IV encompasses the results and investigation. Section V presents the conclusions.

II. Related Work

Declaration and quantification of DNA damage is an actual substantial topic in biological and biomedical study areas, which requires further influential and active methods. Defining the DNA damage level is a significant point to decide the fate of the cell, such as if the cell recovers the DNA damage, or kills itself, or develops into an abnormal cell and forms into a serious diseases such as cancer [7]. Besides, the defining level can help to get more insight over drug treatment experiments. Several attempts have been made to classify DNA damage levels. For instance, numerous classical machine-learning methods have been employed in classifying the data that were related to gene expression, involving Fisher linear discriminant analysis [8], decision tree, k nearest neighbor [9], multi-layer perceptron [10], support vector machine [11] [12], boosting, and self-organizing map [13]. In addition, concerning clustering gene expression data, various machine learning techniques have been utilized [14]; they include hierarchical clustering [15], graph theoretic approaches [16] [17] and self-organizing map [18]. Concerning disease and treatment for DNA damage also, another attempt is available in the literature based on use of the machine learning classifiers on illness datasets, like Leukemia disease dataset, Lymphoma malignance data set and colon tumor dataset. Researchers have also attempted to explore many features by utilizing classical methods, like multi-layer perceptron neural network, k-nearest neighbor, structure adaptive _SOM- self organizing map and SVM (support vector machine); these have been employed for classification [19]. In addition, they have joined the classifiers to increase the performance of classification. The experimental consequences indicate that the ensemble with some basic classifiers produces the greatest classification rate on the benchmark dataset.

Other researchers have established an SVM classifier exactly for mtDNA missense variants [20]. Therefore, in the process which is associated in the training and validation of the model, they employed 2,835 mtDNA damaging and neutral amino acid replacements. In the abovementioned dataset, each instance is well-defined through a fixture of three attributes created on evolutionary preservation in Eukaryote modified amino acids. Consequently, the proposed classifier achieved better than other web-available tested predictors. However, lately, a Deep learning model has been offered [21]. The model is based on a weak label learning method; they used this method to investigate the whole slide images (WSIs) of Hematoxylin besides Eosin (H&E). Their occupation was Self-supervised pre-training technique and heterogeneity aware deep Multiple Instance Learning (DeepSMILE) and they engaged it on cancer tissue images. Their model improvements recommended the genomic label classification performance without collecting larger datasets. There is also a deep learn pipeline based open source, called FociNet [22]. It is interested in image classification and was established to mechanically segment full-field fluorescent images and divide DNA damage of each cell. The outcome from the model indicated that FociNet reached satisfying performance in classification. Since it classifies a solitary cell in a normal, injured, or no signaling (no fusion-protein expression) state, and it also shows exceptional matching in the assessment of DNA damage, contingent on fluorescent foci images from different imaging platforms [23]. Evaluation of the performance of convolutional Neural Network was done to examine the amount of DNA damage by means of comet assay images and was matched to further approaches in the literature. The novelty of their work was employing convolutional Neural Network as a novel scheme to classify the comet objects on segmented comet assay inside the images. Additionally, numerous deep learning models were applied on DNA damage images [24] [25] [26]. However, almost all the available deep learning models in the available literature are based on image datasets for DNA damage, while few available deep learning models are based on DNA damage intensity – concentration datasets; this is due to a scarcity in experimentally observed data concentration datasets. Therefore, while it is challenging to envision these complicated relationships using only DNA damage, investigators can systematically confirm these associations with a mathematical-numerical model that incorporates data from experiments toward a kinematic mathematical model which includes the cell cycle regulation techniques with the DNA damage signaling pathway. Various scientists have developed valuable kinetic mathematical models. These models are associated with the cell cycle regulation mechanism to estimate the exchanges of natural species [1], [7], [27], [28], [29], [30], [31].

A mathematical model was proposed as a benchmark model of the DNA damage-signaling pathway and mimic cell fate selection [30]. The outcome from the novel model was that it offers a dataset for the DNA damage signaling pathway. This dataset exposes the proteins' concentration levels and activities to deal with DNA damage level. For instance, the researchers presented the DNA damage signaling pathway-proteins set-concentrations without DNA damage [30]. These observations from the delivered dataset qualitatively match with biologically appropriate facts. In addition, diverse intensities of DNA damage were found, such as Low-damage, Medium-damage, High-damage, and Excess-damage. These aforementioned DNA damage levels bear a resemblance to actual DNA damage, and are triggered by several values such as 100, 200, 400, and 800 J/m2 doses of UV-irradiation. Further explanation will be clarified in the dataset preparation and analysis section and how we utilized this dataset to train and test the proposed artificial deep learning model to predict the DNA damage level into several classes.

III. Proposed Work and Overview of System Architecture (Attention – Based LSTM) – AT-LSTM Model

Currently, there is no effective computational model, neither a machine learning nor a deep learning model that can be utilized to validate the influence of the intensity-concentration of DNA damage on cell cycle system progression. Consequently, the crucial contributions of this paper essentially comprise the following: we employ the attention model to effectively extract the features of DNA damage big and complex dataset and the LSTM layer in the proposed model performs additive interactions, which can help improve gradient flow over long sequences in training [32]. Matched with classical models, AT-LSTM can

competently maintain and work with non-stationary sequences and detect the nonlinear relationships [33]. Furthermore, compared with deep learning models like RNN, the AT-LSTM can avert the long-term dependence issues and give rise to superior interpretability [34]. The mechanism for the attention in the proposed model makes it simple to recognize how the information in the input sequence influences the final created sequence through the model output process [35]. This might assist in discovering the interior operation mechanism of the model and debug certain precise inputs and outputs. Further, the experimental results on DNA damage datasets determine that AT-LSTM accomplishes more enhanced tasks than standard models.

1) The architecture of the AT-LSTM model: The proposed attention-based LSTM (AT-LSTM) model for DNA damage dataset multiclass prediction comprises two parts: the attention model and the LSTM deep learning model. The attention mechanism is able to adaptively choose the furthermost related input features and provide higher weights to the corresponding original feature sequence. Then and there, we utilize the outcomes of the LSTM deep learning model as input for the attention model to predict the DNA damage level and assign it to several classes.

2) LSTM model: For a stated input raw, $X = (x^1, x^2, \dots, x^n)^T = (x_1, x_2, \dots, x_m) \in R^{(n \times m)}$, n represents the numeral of feature orders -sequences, m stands for the length of the window. $x^k = ({x_1}^k, {x_2}^k, \dots, {x_m}^k)^T \in R^m$ is utilized to denote a sequence (vector) of length m. For biological DNA damage, this sequence can be a protein concentration measurement for the sub-network, which represent the DNA signaling network. We use $x_t = (x_t^1, x_t^2, \dots, x_t^n)^T \in R^n$ to represent a set-group of vectors of n features at time t. Long Short-Term Memory (LSTM) model is declared as follows: Let $x_t, h_t \; and \; C_t$ stand for the input, control state, and the cell state on time step t. Delivering a sequence of inputs $(x_1, x_2, \dots, x_m)$ the LSTM calculates the group of sequence $(h_1, h_2, \dots, h_m)$ and the C-sequence $(C_1, C_2, \dots, C_m)$ as follows:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f \quad (1)$$

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f \quad (2)$$

$$c_t = tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (3)$$

$$C_t = f_t * C_{t-1} + i_t * c_t) \quad (4)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = o_t * \tanh(C_t) \quad (6)$$

such that each equation has a set of special symbols, and identify several functions. For occurrence, σ represents the function of logistic sigmoid, $*$ is a component wise multiplication, and C_t is the weather of the cell that is required to be changed. Also, W_f, W_i, W_C, W_o and b_f, b_i, b_C, b_o are a set of parameters for the model. Besides, these parameters can be learned over the processing. Additionally, f_t, i_t and o_t are likewise christened as a gate for the forgotten, along with an input gate and output gate. In actual fact, the architecture for the LSTM unit includes a memory cell, this mean that every LSTM unit that contains a memory cell has state $\mathbf{C}_t$ at time t, which is structured by the three overhead gates.

3) The attention model: A significant part of human artificial is that it does not directly contract with all feedbacks from the outside world. As a substitute, human's first attention is on the significant sections to acquire the information they require. Correspondingly, the significance of several proteins concentrations in the biological data set is also different, big, and complex, and the other may be critical. It is also essential to emphasize key features first and remove repeated features. Accordingly, with the operative information inspired through the overhead information, we propose an attention model, and this model can apply the optimization part for the input feature sequence in DNA damage level prediction. An attention mechanism [35] can be defined as mapping an enquiry. Moreover, a set of key-value couples to an output, and similarly, the components in the system such as keys, query, values, including output are all defined as vectors. The outcome of the model is calculated as a weighted sum for the values, where the weight given to every value is calculated through a function related to the compatibility for the query with the equivalent key, as shown in Fig. 3.

The method of producing attention weights and the new input features established on attention is illustrated in Fig. 3. In the first fragment, x_t maps to h_t through the following.

$$h_t = f_1(h_{t-1}, x_t) \quad (7)$$

where the non-linear activation function is represented by f_1 , while $h_t \in R^s$ stands for the hidden state on time t, and s indicates the size of the hidden state. LSTM is implemented as f_1 . The main aim for this implementation is to evade the long-term dependence problem, which typically arises in data prediction.

In the second fragment, we generate an attention mechanism by using specifically the deterministic feature in the attention model. For an exact feature sequence like $x^k = (x_1^k, x_2^k, \dots, x_m^k)^T \in R^m$, by relying to the aforementioned hidden state h_{t-1} and the cell state C_{t-1} in the LSTM unit, we express

$$\alpha_t^k = v^T \tanh(W_1 \cdot [\, h_{t-1}, \, C_{t-1}] + W_2 x^k) \quad (8)$$

$$\beta_t^k = softmax(\alpha_t^k) = \frac{\exp(\alpha_t^k)}{\sum_{i=1}^n \alpha_t^k} \quad (9)$$

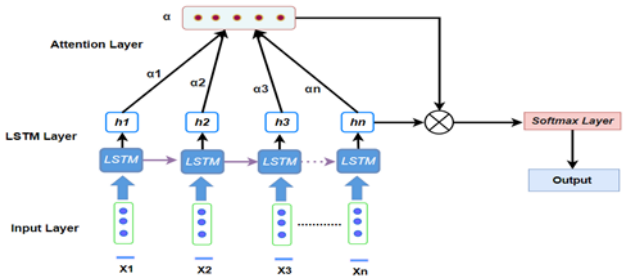


Fig. 3. The Architecture of the Proposed Attention-LSTM Model.

The vector v and the two matrices W_1, W_2 signify the learning abled parameters of the proposed model. The vector a^k has a length called m and its i-th item measures the significance of the k-th input feature sequence at time t. The aforementioned items must be normalized through softmax. β^k represents the weight in attention, which encloses a score, and the score shows the amount of attention that should be put on the k-th feature sequences. We are able to likewise acquire the outcome of the attention model at time t, i.e., the sequence of the weighted input feature named as z_t can be presented as follows:

$$z_t = (\beta_t^1 x_t^1, \beta_t^2 x_t^2, \dots, \beta_t^n x_t^n)^T \quad (10)$$

x_t, in the equations from (1) to (7) swapped via a new calculated z_t to keep up the attention model. However, classical prediction frameworks that enclose recurrent neural networks usually utilized dataset input features as input, besides treating all input feature sequences in an equivalent fashion. Nevertheless, the recently acquired z_t can pay further attention to the particular input feature sequence, mining the key feature sequences efficiently, and based on attention weight, we reduced the influence of the redundant feature sequences. Hypothetically, there would be an improvement in prediction exactness with z_t as the input to the softmax layer.

IV. Results and Discussions

A. Data Analysis and Simulation

This section explores how we apply experiential research on data sets with an aim to elucidate the validity of our DNA damage level prediction framework. First, we will introduce the dataset that was utilized in training and testing of the proposed model. We relied on available biological datasets related to DNA damage signaling pathways, which plays a crucial role in DNA damage in mammalian cell systems [30]. This Biological dataset consists of 15000 records intensity – concentration-level for a five-proteins set which control DNA damage. The DNA damage signaling pathway- proteins set-concentrations without DNA damage was previously presented [30]. These obseravtions from the delivered dataset were qualitatively analyzed with biologically appropriate facts. In addition, diverse intensities of DNA damage were performed, such as Low-damage level, Medium-damage level, High-damage level, and Excess damage level. Herein, we studied and analyzed the aforementioned dataset and proposed a novel model which consists of an attention based on long short term memory- Neural Network named as LSTM (AT-LSTM) model for DNA damage multi-classification prediction.

B. Dataset Analysis

Researchers have assembled a new kinetic based mathematical – ordinary differential equations (ODE's)-model that assimilates the G1/S in cell cycle system models, and they measured compatibility to the biological credibility of the suggested model by confirming numerous mathematical mimicry time progression courses of the intensities of individual biochemical elements [30]. Furthermore, they as well quantitatively recognized the intensity – concentration level of DNA damage and provide experimentally observed data.

Certainly, when DNA damage has occured, numerous protein kinases are involved at the location of damage and launch a special signaling pathway that forces cell-cycle to be arrested. The chief kinase at the damage location is ATM/ATR, which is activated and established on the type of damage and another protein of the gene regulatory protein p53 is also triggered. Mdm2 usually connects to p53 and stimulates its ubiquitylation and destroys the proteasomes. Phosphorylation of p53 stops its binding to Mdm2; consequently, p53 becomes accumulated to maximium levels and inspires transcription of the gene that encrypts the protein p21 and arresting of the cell in G1 [30].

Mainly, in the evolution process for the model, first, we extracted the observation from a base model deprived of DNA damage (DDS = 0) to get the time course for selected cell cycle regulators. Second, we extracted from an expermintal dataset of the benchmark model the required obseravtion with four diverse levels of DNA damage: (Low-damage) with DDS = 0.002, (Medium-damage) with DDS = 0.004, (High-damage) with DDS = 0.008, and (Excess-damage) with DDS = 0.016. If DNA damage has certainly not arisen, the p21 with p53 stop over instead, and with a low level, as illustrated in Fig. 5 and 6. DNA damage drives p53 activation which prompts p21 [2].The character of p21 is to prevent the activity through inhibition of phosphorylation of Rb [1].

With the elimination of DNA-damage, [2]p53 and Mdm2 have a negative feedback loop which is completely reinstated, and p53 returns back to a low-slung level. The reduction in p53 decreases the scale of p21, as shown in Fig. 4 and 5 when DDS=0.004. Likewise, Figure 4 shows all the protein tensity for the protien P21 response in all DNA damage states. For instance, it shows the values for the P21 tensity in the case without DNA damage occurrence as a light-blue line. It provides roughly 3000 instances. On the other hand, it shows the concentration values for the P21 when DNA damage occurrs,; for instance, in case of Low DNA damage, the P21 can be presented in an orange line, medium damage with a gray line, and a high DNA damage level P21 values and behaves to recover the DNA damage represented in a yellow line, while the P21 response in extreme DNA damage is represnted in a dark blue line. We have to be aware about the values for each figure, specifically that each element in each DNA damage state has roughly 3000 different instances in response to DNA damage recovery.

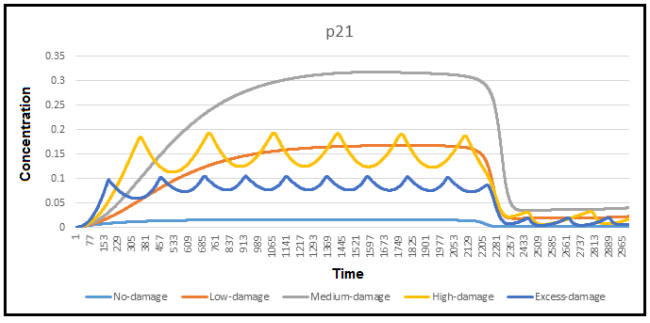


Fig. 4. Time Courses of P21 responses with and without DNA Damage over the Simulation of Mammalian Cell.

With respect to High DNA damage with a rate of DDS=0.008, the time progressions of p21 and p53 are revealed in Fig. 5 and 6. The p53 is activated and presented in oscillation behavior, which were in settlement by means of those previously experimentally detected [30, 6, 29]. Furthermore, when DNA damage is presented, the DNA-damage signal in the sequence triggers p53 instead of Mdm2. The triggered p53 similarly can stimulate the synthesis of p21 which acts as an inhibitor. Meanwhile, p21 stops the phosphorylation of Rb. Fig. 6 indicates all the protein tensity for the protein P53 response in all DNA damage states. For example, it explores the values for the P53 tensity in the case without DNA damage incidence, the light blue line. On the other hand, it displays the concentration values for the P53 when DNA occurred, such as in the Low DNA damage, in which case the P53 is presented with an orange line, and the medium damage is represented with a gray line. In the case of high DNA damage levels, P53 values and behaves to recover the DNA damage that is represented in a yellow line, while the P53 responds in extreme DNA damage cases, represented in a dark blue line.

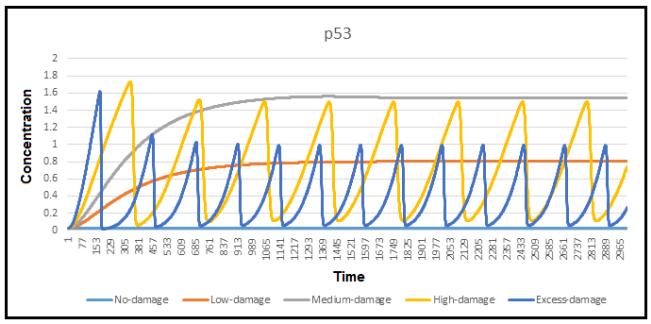


Fig. 5. Time Courses of P53 with and without DNA Damage over the Simulation of Mammalian Cell.

Fig. 6- 8 ilustrate the responses of Mdm2, ATM/ATR and lm respectively in cooprating to handle the DNA damage cases. Each figure, as explained before, shows how the values in protiens tensity of each element will change during the DNA damage, whether it occured or not. As demonstraed before, we have to be alert that the values for each figure of each elelment in each DNA damage state has around 3000 different instances in response to DNA damage recovery.

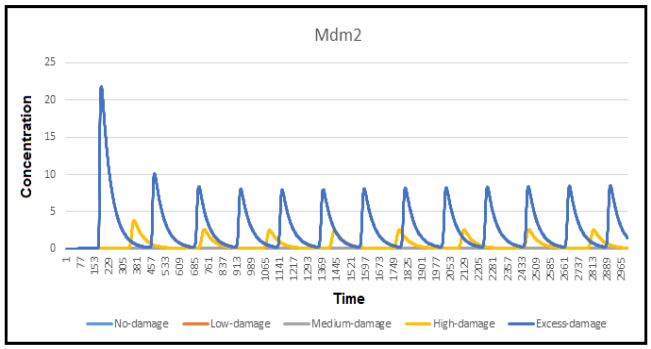


Fig. 6. Time Courses of Mdm2 with and without DNA Damage over the Simulation of Mammalian Cell.

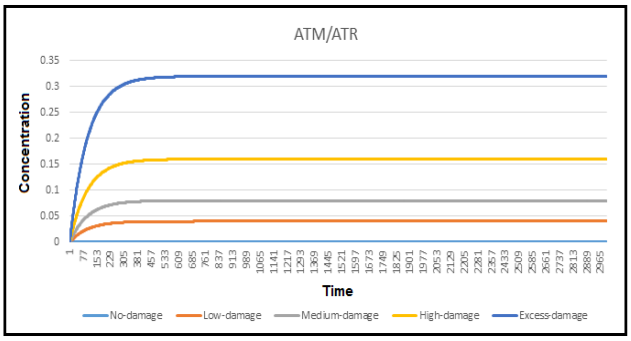


Fig. 7. Time Courses of ATM/ATR with and without DNA Damage over the Simulation of Mammalian Cell.

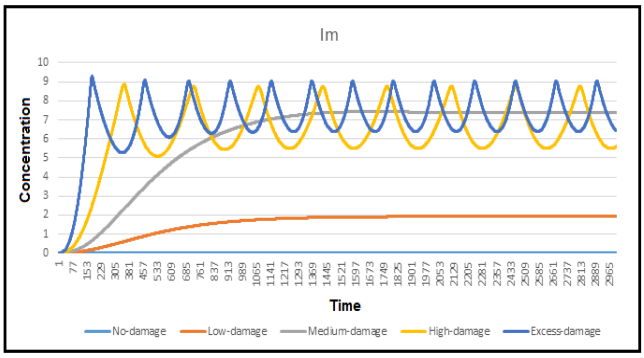


Fig. 8. Time Courses of lm with and without DNA Damage over the Simulation of Mammalian Cell.

Herein, we revealed a brief examination of biological background, specifically cell cycle, with more deliberate focus on DNA damage pathways as a complicated system. As demonstrated before, the novel proposed deep learning Attention based LSTM model is trained and tested depended on the obtained dataset delivered by [30].

C. Experiments

We performed several experiments on the proposed attention-based LSTM model for DNA damage classification. The proposed model was trained on 12000 samples and was tested on 3000 samples. The dataset is in-of-domain for DNA damage classification and the tested dataset that was used is also from the same dataset. There are three cross-validation methods which are often employed to evaluate the success rate of the predictor; namely, the K-fold cross validation, sub-sampling and jackknife test. The Jackknife test is the least arbitrary and most objective, and it has been mostly assumed by researchers to inspect the quality of diverse predictors. This method is source and time consuming. Therefore, in this paper we utlized an early stopping choice to elude a model's overfitting through setting the patience option to three epochs, and we utlized k-fold cross validation where K was set to 1, such that a single train/test split is generated to evaluate the attention-based LSTM model for DNA damage classification.

D. Training

The framework developed by Keras and Python was used to train the attention-based LSTM model for DNA damage classification. For the classification task, SGD optimization algorithm was used with learning rate values set to 0.01 and

momentum set to 0.0, and the model's batch size set to 6. The model initially incorporated 331,525 parameters. The model used 256 LSTM units and the attention dimension was set to 255. Also, the proposed model size proved to require 1.7 seconds per epoch for the classification task. The training data was randomly shuffled at each epoch for the classification task. The proposed attention-based LSTM model for DNA damage classification task was trained to minimize the categorical-crossentropy validation loss for the DNA damage classification task while maximizing validation accuracy for the same classification task.

E. Results Investigation

On performing experiments, it can be assured that the researchers have done extensive experiments on the attention-based LSTM model by testing different hyper-parameters. The proposed model was also experimented using three different configurations, such as BiLSTM with Attention layer, LSTM with Attention Layer and LSTM with Attention and Dropout layers. The classification performances measured will be listed in an accuracy score automatic metric evaluation. The results in Table I show the efficiency of the proposed Attention-Based LSTM model for DNA damage classification. It can be noticed from Table I that the proposed model obtained an excellent result when we used LSTM with the attentional approach, such that the model obtained an accuracy of 93.43. In comparison to other configurations listed in Table I, the proposed model (LSTM with Attention) obtained a better accuracy than the other configurations. These results suggest that the proposed model is effective and accurate in classifying DNA damage in a validation dataset. Also, as seen from Table I, the BiLSTM with attention configurations has obtained a competitive result, such that it obtained an accuracy of 93.13, which indicates that the BiLSTM performs very well in classification tasks. Adding a dropout layer to the model's design negatively impacts the model's performance and quality, such that the model obtained an accuracy of 75.43, as illustrated in Table I. Therefore, it can be summarized that the proposed model outperforms the models that used BiLSTM and Dropout layer. More importantly, results presented in Table I and Fig. 9 show the performance of the model that exploited the attention approach, and LSTM is higher than the other models. In addition, as shown in Fig. 10, it can be seen that the error on the training data decreases as the learning continues, and at the same time, the error of actual valdiation data decreases as the training continues, and this pattern proves that the proposed AT-LSTM model is not facing the problem of overfitting.

Moreover, as shown in Fig. 11, plot of accuracy, we found that the model is trained very well, as the trend for accuracy on both training and test datasets were still rising from epoch 100 till epoch 213, and this is an indication of the proposed model's performance and accurate classification.

Fig. 12. (a,b,c,d,e) illustrates how the AT-LSTM model classifies the responses of P21, P53, Mdm2, ATM/ATR and lm, respectively in cooperating to handle the DNA damage. Each figure shows how the values in protein tensity of each element changes during DNA damage, and how it occurred in each class. The x-axis represents the diverse classes for the DNA damage level while the y-axis represents the amount of each protein concentration during the DNA damage.

TABLE I. EXPERIMENTAL RESULTS

Model Configuration	Accuracy	Number of epochs
BiLSTM +Attention	93.13	228
The proposed model (LSTM with Attention)	93.43	213
LSTM + Attention + Dropout	75.43	71

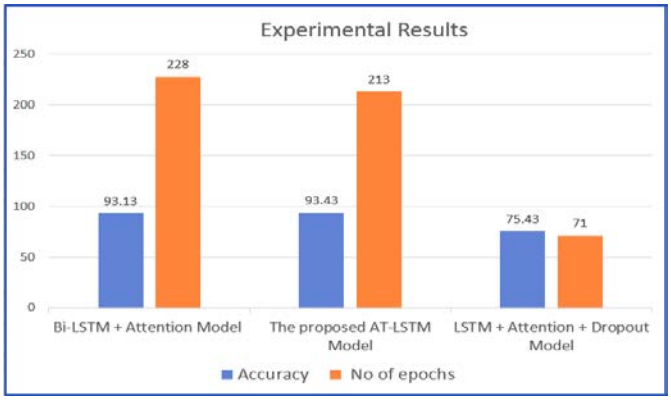


Fig. 9. Model Accuracy within Number of Epochs.

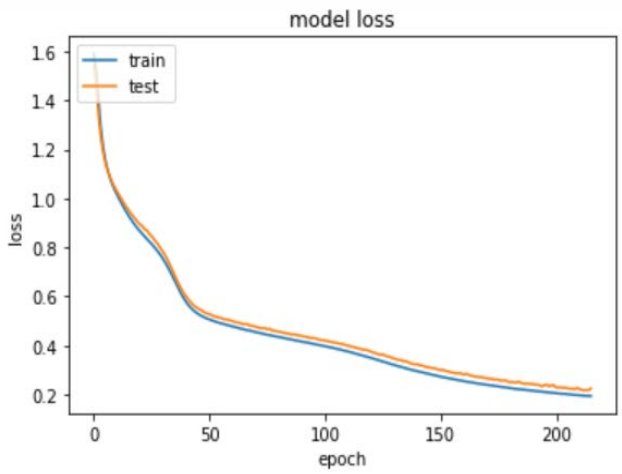


Fig. 10. Model Loss.

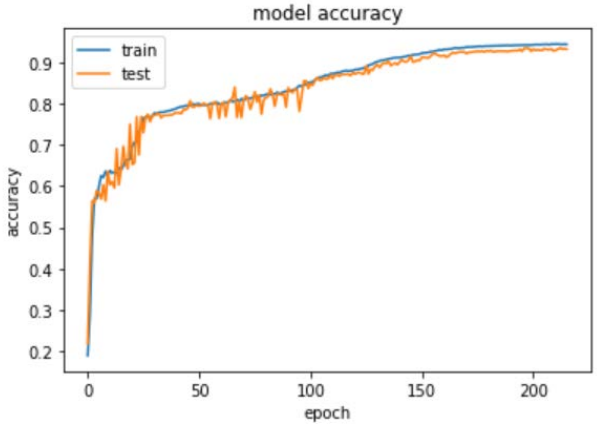
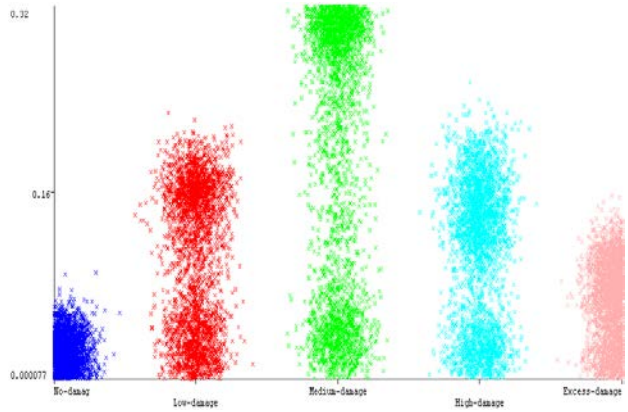
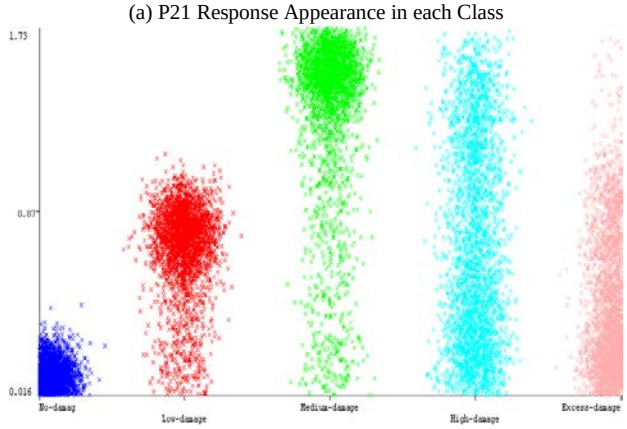


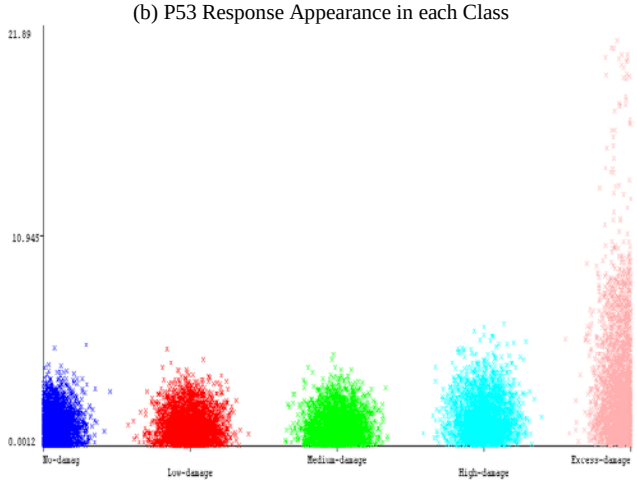
Fig. 11. Model Accuracy.



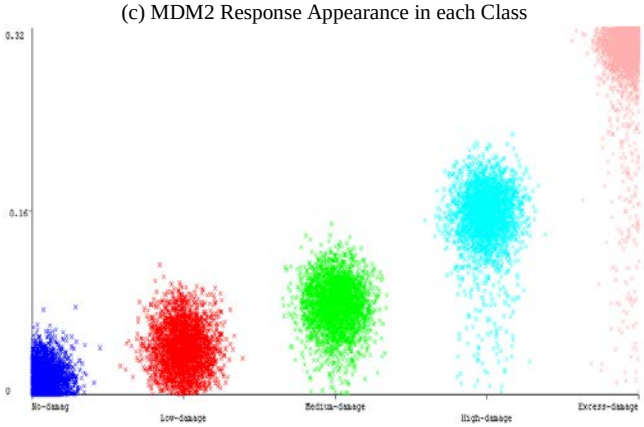
(a) P21 Response Appearance in each Class



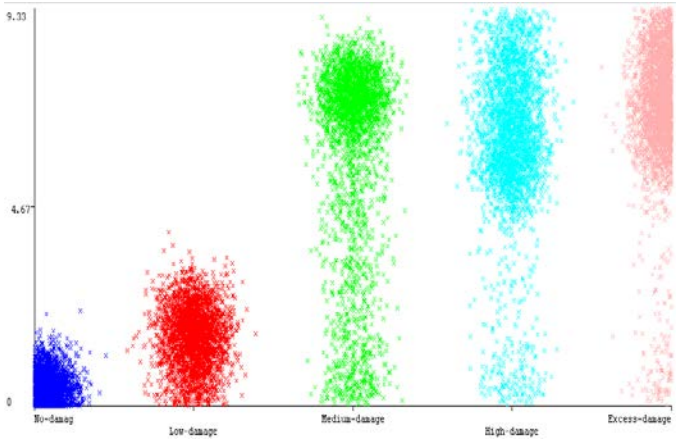
(b) P53 Response Appearance in each Class



(c) MDM2 Response Appearance in each Class



(d) ATM/ATR Response Appearance in each Class



(e) Im Response Appearance in each Class

Fig. 12. AT-LSTM Model Classification for each Protein in Response to DNA Damage Level.

V. Conclusion

DNA damage in mammalian cells causes genetic illnesses and a diversity of cancers. Therefore, more investigation and analysis can help the therapeutic process. Almost all classification prediction models that are used to explore DNA damage are based on DNA damage images, while few studies and models are based on DNA damage intensity. In this paper, we developed a novel deep learning model; in essence an Attention-based LSTM model to perform classification tasks of DNA Damage Levels. The proposed model was able to overcome other models and obtained an accuracy of 93.43%. These results confirm that the proposed model is effective and accurate in predicting and classifying DNA damage on a validation dataset. The attention approach was able to extract the complex features from the dataset and enhanced the proposed model quality and performance. The proposed model is considered as a novel work since AT-LSTM has never been applied in the DNA damage field, and can be employed to assist the investigation and studies of DNA damage since it provided a promising prediction of results.

References

[1] Y. H. H. O. M. H. T. Tashima, "Prediction of key factor controlling G1/S phase in the mammalian cell cycle using system analysis," Biosci., pp. 106 (4), 368–374., 2008.

[2] K. 1. Kohn, "Molecular interaction map of the mammalian cell cycle control and DNA repair systems." Mol. Biol. Cell, pp. 10, 2703–2734, 1999.

[3] G. H. V. Li, "p53-dependent DNA repair and apoptosis respond differently to high- and low-dose ultraviolet radiation," Br. J. Dermatol, pp. 139, 3–10., 1998.

[4] W. K. B. Roos, "DNA damage-induced cell death by apoptosis. Trends," Mol. Med., pp. 12 (9), 440–450., 2006.

[5] N. P. M. Sharbel Weidner Maluf, "DNA Damage and Oxidative Stress in Human Disease," BioMed Research International, 2013.

[6] M. D. Bryant C Nelson, "Implications of DNA damage and DNA repair on human diseases," Mutagenesis, , pp. Volume 35, Issue 1, 2020.

[7] H. T. Y. K. Y. I. K. H. T. E. Y. O. Hamada, "Sophisticated framework between cell cycle arrest and apoptosis induction based on p53 dynamics." PLoS ONE, pp. 8, 28–35, 2009.

[8] S. F. J. a. S. T. P. Dudoit, "Comparison of discrimination methods for the classification of tumors using gene expression data," Department of Statistics, University of California, Berkeley, p. Technical Report 576, 2000.

[9] L. W. C. R. D. T. A. a. P. L. G. Li, "Gene selection for sample classification based on gene expression data: study of sensitivity to choice of parameters of the GA/KNN method." Bioinformatics, pp. 17(12):1131-1142, 2001.

[10] J. W. J. S. R. M. S. L. H. L. M. Khan, "Classification and diagnostic prediction of cancers using gene expression profiling and artificial neura networks," Nature Medicine, pp. 7(6):673-679., 2001.

[11] T. S. C. N. D. N. B. D. W. Furey, "Support vector samples using microarray expression data," Bioinformatics, pp. 16(10):906-914. , 2000.

[12] M. P. S. G. W. N. L. D. C. N. Brown, "Knowledge-based analysis of microarray gene expression data by using support vector machines," Proc. of the Natl. Acad. of Sci. USA, pp. 97:262-267, 2000.

[13] T. R. S. D. K. T. P. H. C. Golub, "Class discovery and class prediction by gene-expression monitoring." Science, pp. 286:531-537, 1999.

[14] R. a. S. R. Shamir, "Algorithmic approaches to clustering gene expression data," Current Topics in Computational Biology, MITpress, 2001.

[15] M. B. S. P. T. B. P. O. a. B. Eisen, "Cluster analysis and display of genome-wide expression patterns," Proc. of the NatlAcad. of Sci. USA,, pp. 95:14863-14868, 1998.

[16] E. S. A. L. J. M.-E. S. Hartuv, "An algorithm for clustering DNA fingerprint," Genomics, pp. 66(3):249-256, 2000.

[17] A. B. L. F. N. N. Ben-Dor, "Tissue classification with gene expression profiles," Journal of Computational Biology , pp. 7:559-584, 2000.

[18] P. H. C. Tamayo, "Molecular classification of cancer: Class discovery and class prediction by gene-expression monitoring," Science, pp. 286:531-537, 1999.

[19] S.-B. C. a. H.-H. Won, "Machine Learning in DNA Microarray Analysis for Cancer Classification," DBLP, 2003.

[20] A. G.-S. Antonio Martín-Navarro, "Machine learning classifier for identification of damaging missense mutations exclusive to human mitochondrial DNA-encoded polypeptides," BMC Bioinformatics, p. Article number: 158 (2017), 2017.

[21] E. G. I. N. H. M. H. J. T. Yoni Schirris, "DeepSMILE: Self-supervised heterogeneity-aware multiple instance learning for DNA damage response defect classification directly from H&E whole-slide images," Electrical Engineering and Systems Science, 2021.

[22] D. X. 1. R. Z. 2. L. Z. Xuechun Chen 1, "Deep-Learning-Assisted Assessment of DNA Damage Based on Foci Images and Its Application in High-Content Screening of Lead Compounds," publmid, pp. :14267-14277., 2020.

[23] Y. B,. E. Ş. Umit Atila, "Classification Of Dna Damages On Segmented Comet Assay Images Using Convolutional Neural Network," Computer Methods and Programs in Biomedicine, p. 186:105192, 2019.

[24] R. H. M. S. a. D. W. G. Christoph Sommer, "A deep learning and novelty detection framework for rapid phenotyping in high-content screening," Molecular Biology of the Cell, pp. Vol. 28, No. 23, 2017.

[25] M. K. Y. S. F. K. O. Jianzhu Ma, "Using deep learning to model the hierarchical structure and function of a cell," Nature Methods, pp. 15, pages290–298 , 2018.

[26] D. M. a. S. Vadivazhagu, "Predicting DNA Damage in Fluorescent Imaging," Worcester Polytechnic Institute, 2020.

[27] Z. W. J. M. W. Qu, "Regulation of the mammalian cell cycle:a model of the G1-to-S transition," Cell Physiol, pp. 284, 349–364, 2003.

[28] B. T. J. Novak, "A model for restriction point control of the mammalian cell cycle," Biol, pp. 230, 563–579, 2004.

[29] A. B. D. C. K. N. B. T. J. Csikasz-Nagy, "Analysis of generic model of eukaryotic cell-cycle regulation.," Biophys, pp. 90 (12), 4361, 2006.

[30] H. H. Y. E. M. O. Kazunari Iwamoto, "Mathematical modeling of cell cycle regulation in response to DNA damage:," BioSystems, p. 384–391, 2011.

[31] H. K. D. S. S. Ling, "Robustness of G1/S checkpoint pathways in cell cycle regulation based on probability of DNA-damaged cells passing," Biosystems, pp. 101, 213–221, 2010.

[32] Z. Fu J, "Look closer to see better: recurrent attention convolutional neural network for fine-grained image recognition," IEEE Conference on Computer Vision and Pattern Recognition, pp. 4476-4484, 2017.

[33] J. L. Elman, "Distributed representations, simple recurrent networks, and grammatical structure," Machine learning, p. 195–225, 1991.

[34] S. N. P. N. Vaswani A, "Attention is all you need," arXiv preprint arXiv:, pp. 1706.03762,, 2017.

[35] C. D. L. H. a. A. M. Yoon Kim, "Structured attention networks," International Conference on Learning Representations, 2017.