Gene mutation detection for breast cancer disease: A review

U N Wisesty¹2, T R Mengko¹ and A Purwarianti¹

¹ School of Electrical and Information Engineering, Bandung Institute of Technology, Bandung, Indonesia
² School of Computing, Telkom University, Bandung, Indonesia
*untarinw@telkomuniversity.ac.id

Abstract. Breast cancer is one of the most common diseases suffered, especially by women, in the world, and about two billion new cases of patients with breast cancer in 2018. Therefore, it is very important to detect cancer early. Early detection of cancer can be done through the analysis of DNA abnormalities from blood cell samples, where the sampling does not require surgery, non-invasive and painless, and can reduce the sampling cost. DNA abnormalities can occur due to heredity or gene mutation. This paper presents a systematic review that includes an explanation of DNA sequences, gene mutations that occur in breast cancer, and bioinformatics techniques for detecting breast cancer. From several studies that have been conducted in the medical field there are mutations in the BRCA1, BRCA2, and PALB2 genes, where mutations in these genes can cause an increased risk of breast cancer. Other gene mutations associated with cancer risk are ATM, BARD1, CDH1, CHEK2, MRE11A, NBN, TP53, PTEN, RAD50, RECQL, RINT1. In bioinformatics, breast cancer detection based on DNA sequence data is carried out in three phases namely data mapping, feature extraction, and prediction / classification. The methods that can be used are Voss mapping and its variations for data mapping, statistical feature representation approach and Wavelet analysis for feature extraction, and regression approaches, probability models, Support Vector Machines, Neural Networks and Deep Learning for classification.

1. Introduction
Cancer is a disease caused by abnormal division of cells that is uncontrolled. One of the most common cancers is breast cancer in women, and an estimated two billion new cases of patients suffering from breast cancer in 2018 [1]. Cancer detection is initially done with tissue tumor biopsy, but this method is very risky for patients. Tissue tumor biopsy also require surgical operations, which can cause pain and risk for patients [2], so, many people prefer other alternatives treatment. In Robertson and Baxter [3], it is stated that not all patients can be analyzed with tissue tumor biopsy because the location of the tumor is hard to reach and not all cancers form solid tumors. Solid biopsy cannot be done many times due to the existing risks and high costs. Whereas in the process of treating cancer, biopsy it needs to be done many times to analyse the development of the cancer. So that in its development, it was found liquid biopsy, which in liquid biopsy uses blood or other biological liquid tissue, painless and non-invasive, does not require surgery, and can minimize costs and time [2].

Early detection of cancer can be done through analysis of DNA abnormalities from the patient's blood sample or other biological fluids. Abnormalities in DNA caused by heredity or gene mutations. In research carried out in the late 19th and early 20th centuries, David von Hansemann and Theodor Boveri...
analyzed the cancer cells division through a microscope and found abnormalities in the structure of their chromosomes [4]. This matter shows the connection between genomic abnormality and cancer. In research [5], it was mentioned that in the case of breast cancer there was a mutation in the patient DNA which caused genome instability related to the mechanism of gene repair. Mutations that occur can be in the form of point mutations, insertions, and deletions. Each type of cancer will cause mutases in certain chromosomes. In the case of breast cancer, the most common mutase occurs in the BRCA1 and BRCA2 genes [6-10] with sample dataset of patients in America, Africa, China, and Europe. Whereas in breast cancer patients from the South Indian population also found mutases in the PALB2 gene [11].

In the field of bioinformatics, several methods have been developed to detect mutations that occur due to cancer. Tuqan developed a method for detecting nucleotide bias in codon structure using the DSP approach [12]. This becomes important for tracking period-3 components using the DNA spectrum. Liu uses four levels of DPT using Haar Wavelet and statistical features to detect gene mutation patterns, and the Support Vector Machine to classify patterns of cancer and non-cancer genes. In that study, it can reach 100% accuracy of the dataset used [13]. Yifeng Li developed an early cancer monitoring system based on the Multi Modal Deep Boltzmann Machine by detecting somatic mutations that can cause changes in normal tissue to become a tumor state [14]. From the experiments conducted, the proposed method can achieve accuracy above the frequency-based method used as the baseline in the paper.

This paper presents a review of methods for detecting gene mutations, specifically those related to breast cancer in the field of bioinformatics. The review included gene mutations related to breast cancer, as well as techniques in bioinformatics for detecting breast cancer which included the method of DNA mapping, feature extraction, and classification. The results of this paper review are expected to be able to assist researchers in developing breast cancer detection systems based on DNA sequences along with machine learning techniques that can be used in these systems.

2. Method

This paper presents a systematic review of breast cancer detection based on DNA sequence data. The review included international journals and proceedings from online databases that included Pubmed Central / National Center for Biotechnology Information (NCBI), Association for Computing Machinery (ACM), Springer Journal, Google Scholar, and Scopus. From the papers and journals that have been obtained, then divided into three parts, namely:

- DNA sequences, including the representation of DNA in nucleotides, mutations that can occur in genes, and DNA sequencing technology.
- Gene mutation in breast cancer, including any gene mutations that can occur for breast cancer.
- Breast cancer detection in bioinformatics, including DNA mapping techniques, feature extraction and prediction / classification methods.

3. Results and discussion

3.1. DNA sequence

Genetic information is stored in cells in the form of nucleic acids and proteins. This information is useful for knowing the structure of living things and the evolutionary processes that occur. DNA is composed of unit structures called nucleotides. Each nucleotide consists of a pentose sugar (2’-deoxy-D-ribose) and one of nitrogenous bases (adenine (A), thymine (T), guanine (G), or cytosine (C)), and a phosphate [15]. Each nucleotide has its own pair, A with T and G and C. In RNA, thymine is referred to as uracil (U). The genetic code can be formed from three nucleotide sequences called codons, which are representations of amino acids. The transcriptional process of codons from DNA to mRNA aims to obtain protein (polypeptide) content in living things. In the process of transcription this begins by looking for the start codon in the DNA sequence. The code for the start codon is ATG in DNA, and AUG in RNA. While the transcription process will stop when a code is found for the stop codon, namely TAG (amber), TGA (opal), and TAA (ochre).
Gene mutations that occur in DNA can be point mutations and base pair mutations. In point mutation, mutase occurs by changing the value of one or several nucleotides, without changing the length of DNA. Point mutations that occur can be silent mutations, missense mutations, and nonsense mutations. Silent or synonymous mutation is a gene mutation that occurs without changing the amino acids produced. For example, the CGC codon mutated into CGA still produces arginine amino acids. In the latest findings, although silent mutations do not change the amino acids produced, but these mutases can change the function of proteins, called altered conformation of the protein, and affect the process of polypeptide folding. Missense or synonymous mutation, the mutase that occurs can change the production of amino acids produced. Whereas nonsense mutation can cause premature stop codons in DNA. Furthermore, base-pair (insertion and deletion) mutation is the addition or deletion of one or more genes in DNA. This causes changes in the length of the DNA itself as well as the amino acids produced.

Genome sequencing is a technology that can be used to detect gene mutation. One method of DNA sequencing is the dideoxy method which was first published by Sanger [16], where the technique used is based on the chain-termination principle. Then, 20 years later, Ronaghi published a new DNA sequencing approach called real-time DNA sequencing based on the results of DNA polymerase activity detection with enzymatic luminometric inorganic pyrophosphate [17]. Today the technique is known as Next-Generation Sequencing (NGS) technology.

3.2. Gene mutation in breast cancer

Cancer is a disease that causes one in eight deaths in the world [18]. This includes more than 100 different types of diseases originating from cell types and organs in the human body, where the cells can attack normal cells and move to other organs that are far away [4]. Cancer is caused by abnormal changes in DNA in cancer cells. However, not all these changes cause more severe cancer. In this case, there is the concept of driver and passenger mutase which can influence the development of cancer cells. Driver mutation will affect the development of cancer cells and is found in the tissue where the cancer is detected. Meanwhile, passenger mutation does not affect the growth of cancer cells but will also be found in the cancer genome as a consequence of the cell division process. Therefore, the analysis of cancer genes aims to identify cancer cells that contain driver mutations and distinguish between driver mutations and passenger mutations.

Some of the most well-known genes cause an increased risk of developing breast cancer, namely BRCA1, BRCA2, and PALB2 [19]. Mutations that occur in these three genes can be inherited on to their children. The function of the BRCA1, BRCA2, and PALB2 genes is to keep breast cells from growing normally and prevent cancer cells from developing. In other studies, [9,20], also studied other genes related to breast cancer risk, including:

- ATM (help repair damaged DNA),
- BARD1 (Working with the BRCA1 gene to repair DNA, mutations in the BARD1 gene can also increase the risk of ovarian cancer),
- CDH1 (producing proteins that can help cells to bind to one another, mutations in this gene can also increase the risk of stomach cancer),
- CHEK2 and p53/TP53 (Instructing the proteins production that can stop tumor growth, mutations can also increase the risk of colon cancer, prostate cancer, Li-Fraumeni syndrome, leukemia, brain tumor, and sarcomas),
- MRE11A, NBN, and RAD50 (form an MRN complex that can help repair DNA damage in cells, mutations in the MRE11A gene are related to ataxia-telangiectasia-like disorder which can affect brain development and weaken the body’s immune system, NBN gene mutations also cause Nijmegen breakage syndrome which can slow growth in the future childhood and other health problems, as well as the RAD50 gene mutation causes the repair process to stop DNA damage),
- PTEN (helps regulate cell growth, PTEN gene mutations can also cause Cowden syndrome),
• RECQL (instructing the RecQ helicases production, which are enzymes that can help repair DNA damage, the mutation of the RECQL gene can also increase the risk of larynx cancer, brain cancer, and pancreatic cancer),

• RINT1 (also referred to as RAD50 interactor 1 which helps regulate cell division, the RINT1 gene mutase is also associated with the risk of Lynch syndrome).

In research, mutation screening studies were carried out in breast cancer cases where the sample used was blood samples from a public hospital in Chile [8]. The number of samples used was 336 patients who had breast cancer or ovarian cancer. The sample was taken from 1999 to 2015, with criteria for three or more families having been diagnosed with breast cancer over the age of 45 years, two families diagnosed with breast cancer before the age of 45 years, one breast cancer patient and ovarian cancer at all ages, one male patient and women suffering from breast cancer. And 117 patients selected for breast cancer and / or ovarian cancer, do not have a family history of breast cancer, and under the age of 40 years. In that study, found 25 mutations (6 novels), where 9 of them only occur in Chilean patients who are not related by family and spread in different areas. Of the 9 mutations found, 4 mutations occurred in the BRCA1 gene and 5 mutations in the BRCA2 gene. Based on the number of patient samples available, 78% mutations occur in the BRCA1 and BRCA2 genes.

Figure 1 shows the mutation position of the BRCA1 and BRCA2 genes. In BRCA1 the gene reading begins with ATG and ends with TGA, whereas in the BRCA2 gene begins with ATG and ends with TAA. BRCA1 538insC mutations are found in native Germans, Poles, and the USA; BRCA1 185dAG was found in native Israelis and the USA; BRCA1 1675delA and 1135insA in native Norwegians; BRCA1 4153delA in native Polish population; BRCA1 3171ins5 in native Sweden; BRCA2 999del5 in native Icelanders; BRCA2 6174del in native Israelis and the USA; and BRCA2 9254del5 in native Spain [21].

Further findings show that if an adult has the BRCA1 and BRCA2 mutations then the chance to be passed on to offspring is 50%. However, the observations made by the inheriting gene mutation process did not occur in patients who were still in childhood. This phenomenon can occur because the mutation is not inherited dominantly and there is intervention from the other parent, so that it can reduce the likelihood that the mutated gene can be inherited [9]. A study was conducted to detect mutations in the PALB2 gene from 2297 patients from Chinese origin and found mutases in 1.31% of patients with a history of breast cancer offspring and 0.56% in other patients. In breast cancer patients from Malaysia, a missense mutation was found in the PALB2 gene with a frequency of 31.1% [22]; North Indian population of 2.9% [23]; South Indian population of 3.5% [11].
3.3. Breast cancer detection in bioinformatics
Cancer is caused by the accumulation of driver mutations and gene abnormalities associated with cancer. The bioinformatics approach is an alternative in detecting the gene mutation. In detecting gene mutations based on DNA sequence data, it is divided into three stages, namely DNA mapping, feature extraction, and prediction / classification.

3.3.1. DNA sequence mapping
In this stage, the conversion of DNA sequence data in the form of strings (nucleotide A, C, T, and G) into numerical representations is carried out. This is necessary because feature extraction and classification methods often require numeric input. The following are methods of converting data representations that can be used on DNA sequence data.

- Voss mapping is a method that widely used to convert DNA sequence data in the form of strings into numeric [24]. The advantage of using this method is very simple and efficient in spectral analysis in DNA sequences. However, the weakness is that Voss Mapping cannot represent the relationship between nucleotides. The method used is a DNA sequence consisting of nucleotides A, T, C, and G, converted into four binary sequences corresponding to each nucleotide, where "1" indicates the existence of certain nucleotides, and "0" if the nucleotide is not appear [12,25].

- Z-Curve is the development of Voss Mapping, where one DNA sequence convert to three vectors (Equation 1, 2, and 3) [24,26]. Nucleotide

\[ x_i = (A_i + G_i) - (C_i + T_i) \]

\[ y_i = (A_i + C_i) - (G_i + T_i) \]

\[ z_i = (A_i + T_i) - (C_i + G_i) \]

- The tetrahedron method aims to reduce the four vectors generated in Voss Mapping to three RGB vectors. Conversion of DNA sequences into tetrahedron representations can be seen in research [26].

- Complex Number Representation aims to represent nucleotide pairs where A with T and G with C, so that the numerical values for each nucleotide can be formulated in Equation 4.

\[ A = -1 + j; T = 1 - j; C = -1 - j; G = 1 + j \]

- Integer and Real Representation, a DNA sequence will be converted into a vector containing integer or real numbers, with integer representation rules \( X(i) = 0 \) for T, \( X(i) = 1 \) for C, \( X(i) = 2 \) for A, \( X(i) = 3 \) for G; and real representations \( X(i) = 1.5 \) for T, \( X(i) = 0.5 \) for C, \( X(i) = -1.5 \) for A, \( X(i) = -0.5 \) for G [26].

- Trigonometric Mapping produces a sequence in which the conversion process uses trigonometric functions, with an angle \( \theta \) of pi / 3 value (Equation 5) [27].

\[ \tilde{X}(i) = \begin{cases} 
-Cos(\theta) + j * Sin(\theta), & X(i) = A \\
-Cos(\theta) - j * Sin(\theta), & X(i) = C \\
-Cos(\theta) + j * Sin(\theta), & X(i) = G \\
-Cos(\theta) - j * Sin(\theta), & X(i) = T 
\end{cases} \]

- Paired Numeric will convert a DNA sequence to a numerical sequence of values 1 and -1 and takes into account the nucleotide pairs A-T and G-C (Equation 6).

\[ \tilde{X}(i) = \begin{cases} 
1, & X(i) = T \cup A \\
-1, & Otherwise 
\end{cases} \]
3.3.2. Feature extraction. The feature extraction phase aims to extract important information from DNA sequence data. The feature extraction method is needed because it can speed up computing time compared to conventional pattern matching methods. Following are the feature extraction methods that can be used in DNA sequence data.

- T. Xu uses statistical features based on DNA sequence data to detect noncoding mutations. These statistical features include [28]:
  a. Tfbs_cnt, tfbs_{max|avg}_sc: the number of site bindings of overlapping transcription factors, along with the maximum number and average.
  b. <tfname>_cnt, <tfname>_{max|avg}_sc: maximum value, sum and average value of the binding site factor transcription that has the name of a gene.
  c. Dhs_src_cnt: the amount of DHS (DNase I hypersensitive site) containing the mutation.
  d. Dhs_max_sc: the maximum value of DHS containing the mutation.
  e. Gerp_sc: Gerp conservation score.
  f. Tss_dist: base-pair distance to the nearest TFBS.
  g. Gc_per: presentation of GC-content.

When using these features, the available data must be already having an annotation where the mutase occurred.

- Wavelet Analysis aims to extract important features from DNA sequences that have been converted into numerical vectors. The wavelet method was originally used in signal processing, which uses mathematics based. One of wavelet method is Discrete Wavelet transform (DWT), which processes signals using a low pass filter and decomposed it using a high pass filter [28].

3.3.3. Gene prediction/classification. In bioinformatics, gene prediction or classification is an approach that can be used for gene annotation, gene finding [29], gene mutation detection, and detecting whether a person has cancer or not. The methods used also vary, starting from string matching, statistical analysis, or by machine learning approaches. In detecting gene abnormalities associated with early detection of cancer, the introns and exons are separated first because the abnormalities that cause cancer occur in the exons area. Zhang proposed the Multiscale Products of Multiscale Bilateral Filtering (MP-MSBF) method to predict exon areas. In this study, DNA sequence data is converted into numerical numbers using paired-numerical representation, then perform MSBF calculations, multiscale products, Jensen-Shannon (JS) divergence, and inter-scale correlation analysis [30]. In this study, it was concluded that the exons area has a smaller JS divergence than the area of the introns, and the proposed method is non-linear filtering so that it can predict exons that have a smaller area.

Meanwhile, Das proposed a DNA sequence mapping method called trigonometric mapping to improve the prediction of the exons area using a digital filtering method [27]. One of the data used in the study is the Homo Sapiens Tubulin gene originating from NCBI, and performance measurements are carried out by comparing the program output with annotations conducted by NCBI. The performance obtained from this research is better than other mapping methods, namely representation of integers, EIIP, DNA Walk, paired numeric, and entropy. Table 1 shows a review of each approach in gene prediction or classification.
Table 1. Review of gene prediction / classification research in bioinformatics.

| Gene Prediction Approach                                                                 | Application in DNA Sequence Analysis | Review                                                                                     |
|-----------------------------------------------------------------------------------------|--------------------------------------|-------------------------------------------------------------------------------------------|
| Statistical feature extraction, Generalized Linear Regression Models based on Poisson distribution, Ensemble Decision Tree based on Random Forest, AdaBoost, Gradient Boosting Regression Tree (GBRT), Feed Forward Neural Network (FFNN), and Stack Autoencoder (SAE) [28]. | Noncoding Annotation Mutation        | In that study, Poisson regression has the best performance, and FFNN and SAE have performance close to Poisson Regression. In this case, SAE / deep learning requires more training data to achieve better performance. Whereas AdaBoost has the worst performance of all the methods used and does not have predictive power. DNA sequence data is first converted into numerical vectors by replacing A, T, G, and C with 1, -1, j, and -j. Then, the vector is processed using DWT, statistical feature extraction which includes mean, median, standard deviation, range, mean absolute deviation, and mean absolute deviation. By using SVM as a classification method of characteristic statistic features, the study can achieve 100% accuracy for lung, ovarian and breast cancer data. |
| Numerical mapping, Discrete Wavelet Transform (DWT), statistical properties, Support Vector Machine (SVM) [13]. | Cancerous and non-cancerous gene classification. |                                                                                                                                                       |
| BSSV (Bayesian based Somatic Structure Variation) [31].                                  | Somatic Structure Identification of Breast Cancer. | The research aims to detect changes in somatic structure (gene mutation) which include deletion, insertion, and inversion, in breast cancer patients. In this study, it can reach an average precision of 0.9 and a recall of 0.9. These results are proven to achieve higher performance than other published tools namely Breakdancer, GASVPro, and PeSV-Fisher. |
| Expectation Maximation (EM) algorithm, Tumor Covariate Signature Model (TCSM) [32].      | Mutation detection of breast cancer.    | The study detects gene mutations in the Breast Cancer gene data. The EM algorithm is used to optimize the parameters value in the TCSM model. From the experiments conducted, TCSM model can obtain higher performance than the baseline method used (NMF and TCSM using no covariate), using simulation data and original data. |
| Network based method to infer cancer progression (NetInf) [33].                          | Mutation Detection.                   | NetInf algorithm does not require biological knowledge. When compared with the previous method, the NetInf algorithm has several advantages, namely smaller computational complexity by constructing gene networks based on each gene pair, does not require the number of paths in the progression model, and the algorithm infects progression cancer at the pathway level. |
Table 1. Cont

| Method | Description | Detection | Method Aim | Detection Result |
|--------|-------------|-----------|------------|-----------------|
| Multi-modal deep Boltzmann Machine [14]. | Mutation detection of breast cancer related gene. | The proposed method aims to detect mutation that occur in genes associated with breast cancer. The study also compared with other machine learning methods, namely partial least squares and discriminative models. However, both methods cannot detect the existing mutation caused by many irrelevant mutations and sparse data. By using the multi-modal DBM method, 36 mutations (out of 38 genes) can be detected related to the growth of cancer cells. | Mutation detection of breast cancer related gene. | The proposed method aims to detect mutation that occur in genes associated with breast cancer. The study also compared with other machine learning methods, namely partial least squares and discriminative models. However, both methods cannot detect the existing mutation caused by many irrelevant mutations and sparse data. By using the multi-modal DBM method, 36 mutations (out of 38 genes) can be detected related to the growth of cancer cells. |
| SNVHMM based on discrete hidden markov model [34]. | Mutation detection of breast tumor. | The research aims to detect Single Nucleotide Variants (SNVs) which can show gene mutations in cancer cells. The study uses data with low sequencing depth, where if the detection process is carried out with the previous algorithm it still has low performance. The author also conducted experiments using SNVMix, and it was proven that SNVHMM had a higher performance with an f-score of 0.85. SNVHMM can reduce false positives and increase true negatives and requires less training data compared to SNVMix. | Mutation detection of breast tumor. | The research aims to detect Single Nucleotide Variants (SNVs) which can show gene mutations in cancer cells. The study uses data with low sequencing depth, where if the detection process is carried out with the previous algorithm it still has low performance. The author also conducted experiments using SNVMix, and it was proven that SNVHMM had a higher performance with an f-score of 0.85. SNVHMM can reduce false positives and increase true negatives and requires less training data compared to SNVMix. |

4. Conclusion

Cancer early detection becomes very necessary along with the increasing number of new cases and the number of deaths caused by cancer each year. Early detection of cancer can be done by conducting a DNA test from a patient's blood sample. While solid tumor biopsy is relatively difficult to do if the cancer does not form a solid tumor or the location of the cancer-infected organ is difficult to reach. Cancer is characterized by abnormalities in DNA that can be caused by cancer-related gene mutation or inheritance. In breast cancer, gene mutation often occurs in the BRCA1, BRCA2, and PALB2 genes. In bioinformatics, breast cancer detection based on DNA sequence data is carried out in three phases namely data mapping, feature extraction, and prediction / classification. Data mapping aims to change the data representation from string to numeric. The feature extraction phase is needed to take important features in the sequence. Data mapping can be done using Voss mapping and its variations, and feature extraction using statistical feature representation approach and wavelet analysis. The gene prediction / classification phase aims to carry out gene annotation, gene finding, gene mutation detection, and detect whether someone has cancer or not. The methods that can be used are starting from regression approaches, probability models (Bayesian and Hidden Markov Models), Support Vector Machines, Neural Networks and Deep Learning. Research opportunities that can be done are in terms of DNA data representation that can capture the relationship between nucleotide and feature extraction because DNA sequence data have a very large dimension and still in the form of string sequence. Also, research on
breast cancer detection based on DNA sequence data from Indonesian patients in the field of bioinformatics is still very minimal due to the data availability. From the problems that have been explained above, it is expected that research on breast cancer detection based on DNA sequence data from Indonesian patients can be developed.

Acknowledgment
The author thanks the School of Electrical and Information Engineering, Bandung Institute of Technology, for supporting this research.

References
[1] Gorgannezhad L, Umer M, Islam M N, Nguyen N T and Shiddiky M J 2018 Circulating tumor DNA and liquid biopsy: opportunities, challenges, and recent advances in detection technologies Lab on a Chip 18(8) 1174-1196
[2] Marrugo-Ramírez J, Mir M and Samitier J 2018 Blood-based cancer biomarkers in liquid biopsy: a promising non-invasive alternative to tissue biopsy International journal of molecular sciences 19(10) 2877
[3] Robertson E G and Baxter G 2011 Tumour seeding following percutaneous needle biopsy: the real story! Clinical radiology 66(11) 1007-1014
[4] Stratton M R, Campbell P J and Futreal P A 2009 The cancer genome Nature 458(7239) 719-724
[5] Jaksik R and Fujarewicz K 2018 Detection of Genetic Aberrations in Cancer Driving Signaling Pathways Based on Joint Analysis of Heterogeneous Genomics Data In International Conference on Information Theoretic Security 484-494
[6] Dean M, Boland J, Yeager M, Im K M, Garland L, Rodriguez-Herrera M and Lee H J 2015 Addressing health disparities in Hispanic breast cancer: accurate and inexpensive sequencing of BRCA1 and BRCA2 Gigascience 4(1) s13742-015
[7] Lynch H T, Casey M J, Snyder C L, Bewtra C, Lynch J F, Butts M and Godwin A K 2009 Hereditary ovarian carcinoma: heterogeneity, molecular genetics, pathology, and management Molecular oncology 3(2) 97-137
[8] Alvarez C, Tapia T, Perez-Moreno E, Gajardo-Meneses P, Ruiz C, Rios M and Carvajal-Carmona L 2017 BRCA1 and BRCA2 founder mutations account for 78% of germline carriers among hereditary breast cancer families in Chile Oncotarget 8(43) 74233
[9] Lambrou G I, Barbounaki I, Tzortzatou-Stathopoulou F, Petropoulou O, Katrakazas P, Iliopoulou D and Koutsouris D D 2017 Computational Analysis of BRCA1 Mutations in Pediatric Patients with Malignancies and Their Mothers In 2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS) 138-143
[10] Mann G J, Thorne H, Balleine R L, Butow P N, Clarke C L, Edkins E and Giles G G 2006 Analysis of cancer risk and BRCA1 and BRCA2 mutation prevalence in the kConFab familial breast cancer resource Breast Cancer Research 8(1) R12
[11] Kumar H V, Elancheran M, Dhamotharan S R and Indrani J C 2019 Novel PALB2 deleterious mutations in breast cancer patients from South Indian population Gene Reports 17 100492
[12] Tuqan J and Rushdi A 200 A DSP approach for finding the codon bias in DNA sequences IEEE Journal of Selected Topics in Signal Processing 2(3) 343-356
[13] Liu D W, Jia R P, Wang C F, Arunkumar N, Narasimhan K, Udayakumar M and Elamaran V. 2019 Automated detection of cancerous genomic sequences using genomic signal processing and machine learning Future Generation Computer Systems 98 233-237
[14] Li Y, Fauteux F, Zou J, Nantel A and Pan Y 2019 Personalized prediction of genes with tumor-causing somatic mutations based on multi-modal deep Boltzmann machine Neurocomputing 324 51-62
[15] Choudhuri S 2014 Bioinformatics for beginners: genes, genomes, molecular evolution, databases and analytical tools (Elsevier)
[16] Sanger F, Nicklen S and Coulson A R 1977 Real-Time DNA sequencing with chain-terminating
inhibitors Proc. Natl. Acad. Sci. 74(12) 5463–5467
[17] Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlén, M., & Nyrén, P. (1996). Real-time DNA sequencing using detection of pyrophosphate release. Analytical biochemistry, 242(1), 84-89.
[18] Garcia M, Jemal A W E C M H Y S R T M J, Ward E M, Center M M, Hao Y, Siegel R L and Thun M J 2007 Global cancer facts & figures 2007 (Atlanta, GA: American cancer society)
[19] BreastCancer.org Genetic Testing [Online] retrieved from https://www.breastcancer.org/symptoms/testing/genetic.
[20] BreastCancer.org, Other Abnormal Gene Testing [Online] retrieved from https://www.breastcancer.org/symptoms/diagnosis/other-gene-testing
[21] Fackenthal J D and Olopade O I 2007 Breast cancer risk associated with BRCA1 and BRCA2 in diverse populations Nature Reviews Cancer 7(12) 937-948
[22] Phuah S Y, Lee S Y, Kang P, Kang I N, Yoon S Y, Thong M K and Teo S H 2013 Prevalence of PALB2 mutations in breast cancer patients in multi-ethnic Asian population in Malaysia and Singapore PLoS One 8(8)
[23] Singh J, Thota N, Singh S, Padhi S, Mohan P, Deshwal S and Ahmed R 2018 Screening of over 1000 Indian patients with breast and/or ovarian cancer with a multi-gene panel: prevalence of BRCA1/2 and non-BRCA mutations Breast cancer research and treatment 170(1) 189-196
[24] Meng T, Soliman A T, Shyu M L, Yang Y, Chen S C, Iyengar S S and Iyengar P 2013 Wavelet analysis in current cancer genome research: a survey IEEE/ACM transactions on computational biology and bioinformatics 10(6) 1442-14359
[25] Akhtar M, Epps J and Ambikairajah E 2007 On DNA numerical representations for period-3 based exon prediction In 2007 IEEE international workshop on genomic signal processing and statistics 1-4
[26] Mendizabal-Ruiz G, Roman-Godinez I, Torres-Ramos S, Salido-Ruiz R A and Morales J A 2017 On DNA numerical representations for genomic similarity computation PloS one, 12(3)
[27] Das L, Nanda S and Das J K 2017 A novel DNA mapping scheme for improved exon prediction using digital filters In 2017 2nd International Conference on Man and Machine Interfacing (MAMI) 1-6
[28] Xu T and Ramsey S A 2016 Exploration of regression models for cancer noncoding mutation recurrence In Proceedings of the 7th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics 546-553
[29] Stanke M and Waack S 2003 Gene prediction with a hidden Markov model and a new intron submodel Bioinformatics 19(2) ii215-ii225
[30] Zhang X and Pan W 2019 Exon prediction based on multiscale products of a genomic-inspired multiscale bilateral filtering PloS one 14(3)
[31] Chen X, Shi X, Shajahan A N, Hilakivi-Clarke L, Clarke R and Xuan J 2014 BSSV: Bayesian based somatic structural variation identification with whole genome DNA-seq data In 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society 3937-3940
[32] Robinson W, Sharan R and Leiserson M D 2019 Modeling clinical and molecular covariates of mutational process activity in cancer Bioinformatics 35(14) i492-i500
[33] Wu H, Gao L and Kasabov N K 2016 Network-based method for inferring cancer progression at the pathway level from cross-sectional mutation data IEEE/ACM transactions on computational biology and bioinformatics 13(6) 1036-1044
[34] Bian J, Liu C, Wang H, Xing J, Kachroo P and Zhou X 2013 SNVHMM: predicting single nucleotide variants from next generation sequencing BMC bioinformatics 14(1) 225
[35] Roy S S and Barman S 2018 A non-invasive cancer gene detection technique using FLANN based adaptive filter Microsystem Technologies 1-16