Biomedica
Simple & Precise

Genomics Analysis Platform - Muscular Dystrophy Assist powered by Intel® Movidius™ and OpenVINO™ toolkit

Biomdcare uses Avalue's server based on the Intel® Xeon® Processors and OpenVINO™ toolkit to accelerate Muscular Dystrophy Analysis.

Executive Summary

About Biomdcare Corporation

Biomdcare believes “Simple and Precise“ will be the new keystone to drive the next generation medical science industry, and quantifying the human body as well as analysing disease has always been the mission of us all. Established in 2016, Biomdcare makes good use of the high-quality medical service and technology in Taiwan to integrate know-how and experts from different domains to build our exceptional solutions, platforms and tools for clients. Biomdcare empowers doctors and medical researchers with simple yet precise solutions to overcome the challenge of disease screening by leveraging our expertise in medical AI.

Using the next generation of medical technology that integrates artificial intelligence into medicine, we provide medical professionals with simple and accurate solutions to overcome the challenges of screening for diseases. Our solutions and services are based on medical imaging and molecular medicine, including AI screening tools for osteoporosis, cervical cancer, HPV, blood cancers, and other diseases, as well as the X1 series of software, which extends to the PatientSense system for efficient case management and DataSense, a high-performance workstation for easy management of large volumes of medical data.

About Avalue Technology

Founded in 2000 and with headquarters in Taiwan, Avalue Technology (TAIEX: 3479-TW) is a professional industrial computer manufacturing company, who is dedicated to developing the x86 architecture products, including Industrial & Embedded Motherboard, Industrial Computer, Panel PC, System On Module, POS Terminal, Tablet, Embedded Software and various solution ready products. Avalue is dedicated to designing and manufacturing a broad range of IoT solutions included Smart Healthcare, Smart manufacturing and Intelligent Transportation that live up to customers’ expectations, and contribute to a more convenient living environment.

Having expanded, Avalue offers its expertise on PCB/ Assembly/ BIOS version control and all types of after-sales services. An ISO 9001:2015, ISO 13485:2016, ISO 14001:2015 and ISO-45001: 2018 certified company; Avalue offers assurance to customers in every aspect of business. With headquarter located in Taiwan, Avalue has global subsidiaries, including offices in Shanghai, New Jersey, California and Tokyo. In addition, Avalue Technology operates an extensive distribution network to accommodate and serve customers all around the world.

This white paper highlights three use cases of Biomdcare's solutions:

The first is Biomedica X1 Sequencing, Biomdcare's NGS AI-assisted screening software. This tool helps to filter through a large amount of data for doctors to focus on their precious time on the data that cannot be interpreted by machine to expedite the diagnosing process.

The second is Biomedica PatientSense, Biomdcare's laboratory information system. Focusing on patient data and treatment course, integrating various data such as medical records, examination results, survey reports, and etc. Furthermore, help doctors establish a NGS WES analysis database for future visualisation.

The third is Biomedica DataSense, Biomdcare's AI workstation. In the field of biology and medical research, a large number of huge and precious data will be produced.
The management and storing of these data is a major problem we face. DataSense acts as the core of the lab that helps it integrate all the workflows and data into an easy and manageable platform.

Current Challenge

Muscular dystrophies are a group of muscle diseases caused by mutations of genes. As time goes by, muscle weakness decreases its mobility. There are many different kinds of muscular dystrophy, each category affects specific muscle groups. Thus, with the symptoms appearing by ages, it is never irreversible. There are maybe numeral different genetic types in each kind of muscular dystrophy, and people with the same kind of muscular dystrophy can still suffer from different symptoms.

The muscular dystrophies are inherited gene disorders with progressive muscle wasting and weakness of variable distribution and severity. Mutations in the dystrophin gene lead to Muscular dystrophies are rare, and only with few data on how many people are affected.

In inherited muscular dystrophy, the faulty gene is passed down from both parents, neither of whom will have symptoms of the disease. But their children will have a 25% chance of developing muscular dystrophy. The best solution will be restricting this genetic disorder to this generation.

The Whole-exome sequencing (WES) has become an accurate diagnostic test for patients, but it requires 24 weeks. WES has become a standard method for detecting genetic variants in human diseases. Its huge amount of produced data is approximate 10 GB, and it requires revolutionary IT skills and computational power.

Solution

The healthcare industry requires a fast and accurate tool for the genetic analysis

Compared to first-generation sequencing technologies, because of its high throughput and low- cost NGS can rapidly sequence a large number of individual genes and then validate them through clinical databases. It is expected to identify the disease susceptibility genes and related variants in a clinical database, which can be used to predict or suggest the risk of disease or provide relevant information for follow-up treatment. The information has become an indispensable tool to promote precision medicine. In this case, our muscular dystrophy screening kit can shorten this process to 4 weeks.

The amount of data generated by NGS is very large, which is beyond the processing capacity of personal computers or workstations in general laboratories. Most 2 of the mainstream information analysis tools used are developed by academics, but academics usually focus on publishing papers and do not pay special attention to the convenience and optimisation of use and the ability to interface with other tools, so the barrier to entry is relatively high.

NGS application

Due to the promotion of precision medicine or genomic projects in many countries and the increasingly low cost of gene sequencing. It becomes an important use of the huge database of genetic data, to facilitate the analysis and comparison of mutant points for subsequent sequencing. But there is still no standard for this application and requirement for bioinformatics. In clinical applications, the doctors are responsible for the interpretation.

With the approval of the US FDA or NGS-related diagnostic products, it means a new business model for NGS genetic testing industry. With the advantage of being closely related to clinical diagnosis, NGS can not only provide doctors with more accurate medical decision information, but also establish a reasonable medical insurance coverage mechanism and expand the user group; after accumulating a large amount of genetic data, it can increase the opportunity to cooperate with international pharmaceutical companies for new medicine development.

Whole dystrophin gene sequencing by next-generation sequencing (NGS) may be a useful tool for the genetic diagnosis of muscular dystrophies.

This test is recommended for people with a family history of spinal muscular atrophy who are expecting a child or are pregnant. Or couples with a confirmed diagnosis of muscular dystrophy or who have given birth to a child with the disorder can undergo prenatal fetus diagnosis after 10 weeks of pregnancy. And for the general public, they can also undergo this test at their own expense if they are worried about giving birth to a child with the disease.

Innovation: Integrating self-developed AI model with OpenVINO™ toolkit

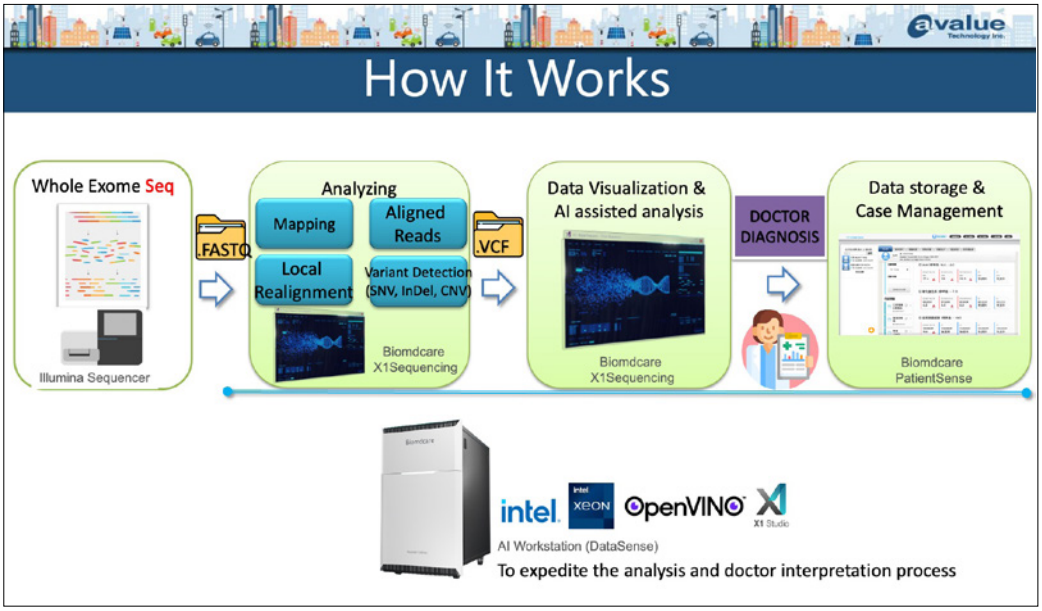
Biomdcare has developed a system that uses deep learning to improve the efficiency of genetic diagnosis for muscular dystrophy.

The system is designed to assist doctors in the diagnostic process by feeding a convolutional neural network (CNN) with gene expression heat maps obtained from the output of a next-generation sequencer (NGS) to infer various patterns of abnormal gene occurrence that are thought to cause muscular dystrophy.

As mentioned previously, the system consists of the "Biomedica X1 Sequencing" application, which screens NGS output, "Biomedica PatientSense," which centrally manages the test information, and the "Biomedica DataSense" AI workstation, which executes inference processing of X1 Sequencing at high speed. Here, we provide an overview of Biomedica X1 Sequencing and Biomedica DataSense.

The general flow of data in the system is shown in Figure 1. First, Whole Exome Sequencing (WES) is initially run on the Illumina NGS sequencer. Next, the NGS output file (.FASTQ format) is fed to the SoftGenetics NextGENe software to run various types of analysis such as read alignment and variant detection. That output is visualized and analyzed with X1 Sequencing and presented to the doctor. The analysis results and doctor interpretation, etc. are stored in PatientSense.

Figure 1. Flow of data within the genetic diagnosis system for muscular dystrophy



The X1 Sequencing application developed by Biomdcare is a key part of this system.
That configuration is shown in Figure 2. This architecture can be applied not only to muscular dystrophy, but also to various diseases caused by gene mutation.

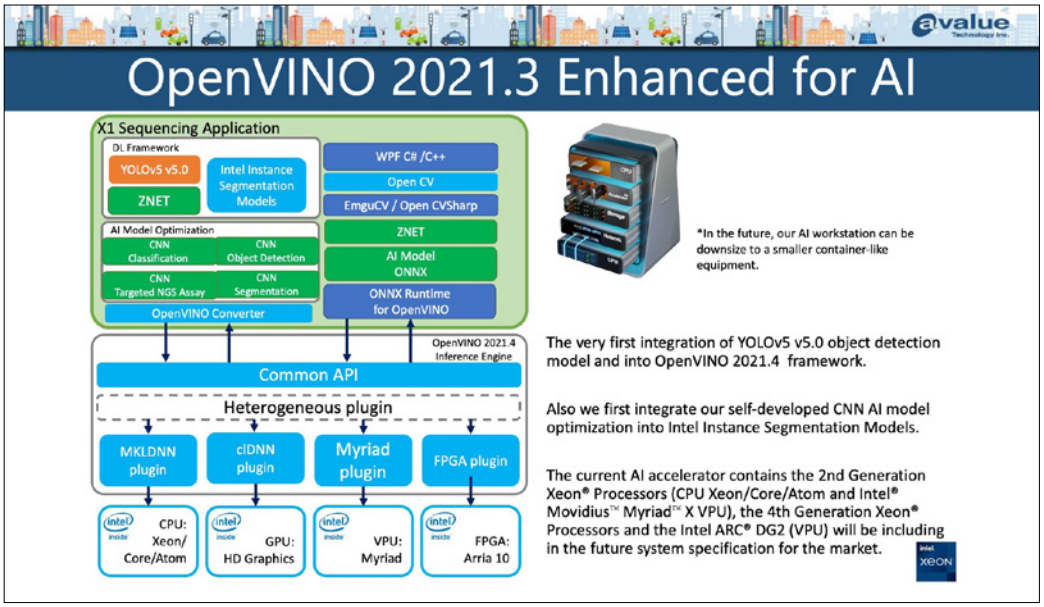
Its features include the adaptation of "YOLOv5," which excels at high-speed object detection, to be compatible with ONNX, making it tractable within the OpenVINO™ toolkit environment, and the integration of the CNN AI model self-developed by Biomdcare with the "Instance Segmentation Model" bundled with the OpenVINO™ toolkit 2021.4 for instance recognition and localization.

Due to restrictions in the 2021.4 release of the OpenVINO™ toolkit such as the inability to directly handle YOLOv5 .pt files, the inference model is converted to ONNX format and fed into the OpenVINO™ toolkit to generate execution files for Intel processors. (There are other methods for generating execution files from ONNX, but the OpenVINO™ toolkit was adopted for its high compatibility with Intel® processors.)

The "Execution Provider for ONNX Runtime" provided by the OpenVINO™ toolkit was used to generate the execution file. Furthermore, because there are some restrictions in the Execution Provider for ONNX Runtime, the ZNET feature for communicating with Intel processors that cannot be supported by the Common API was implemented to mitigate such restrictions.

Moreover, to resolve the image processing problems in AI training, interactions with the OpenVINO™ toolkit Common API are streamlined through ZNET, developed by Biomdcare, in addition to the combined use of the OpenCV library developed by Intel.

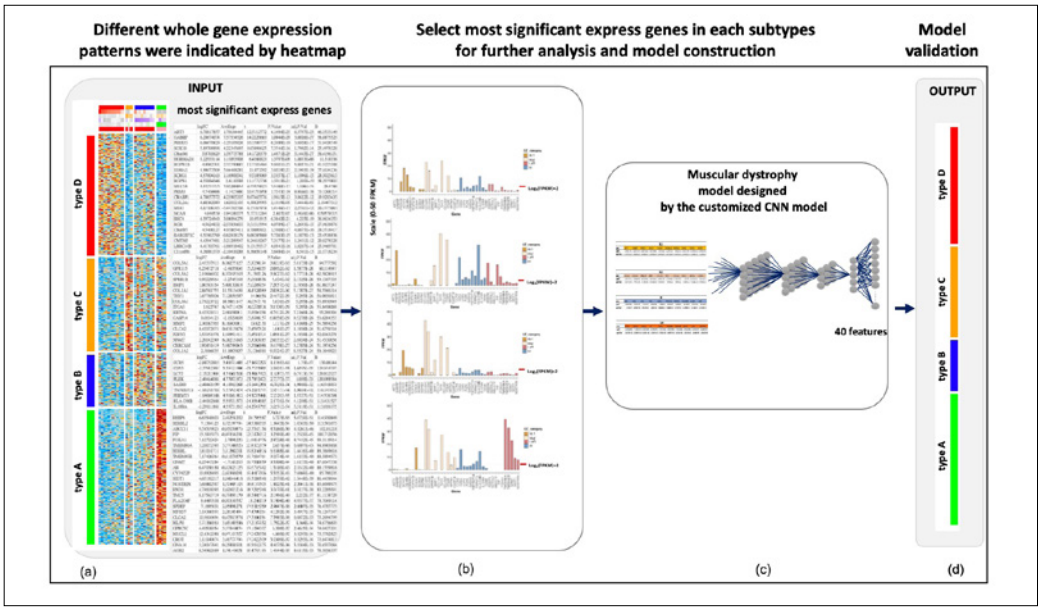
Figure 2. Overview of X1 Sequencing



A representation of the gene expression inference model learning is shown in Figure 3.
Gene expression heat maps created through cluster analysis are treated as images to train the model on features such as the location of target genes and the magnitude of the manifestation frequency.

Data provided by Biomedica's partner hospitals and publicly available open data were used as the genetic data required for learning. Due to a lack of genetic data on muscular dystrophy patients, there is a tendency for overfitting to occur. Biomdcare developed its own method called "Low-resource Training" which enables highly accurate learning even with small amounts of data.

Figure 3. Overview of AI training on gene expression patterns



Innovation: Best workstation for lab researchers and medical professionals

In the field of biology and medical research, a large number of huge and precious data will be produced, and the management and storing of these data is a major problem we face. Biomedica DataSense is designed for laboratories with flexibility, and is capable to support the data science computing requirements.

DataSense provides computerised solutions for high-throughput next-generation gene sequencing data analysis, advanced biostatistics, data exploration, and big data integration. The facility is equipped with servers to handle large amounts of data and powerful computing functions, and to build and install analysis platforms, analysis processes, and tools.

The current DataSense spec consists of 8 (maximum of 20) second-generation Intel® Xeon® Silver 4210 processors (13.75M cache, 2.20 GHz), maximum main memory of 768 GB, an IEI Integration Mustang-V100-MX8 accelerator card equipped with an Intel® Movidius™ Myriad X vision processing unit (VPU), a 10-Gbps x 12-port network switch, and an NVMe SSD (optional) with a maximum speed of 3.2 GB/s.

Furthermore, the cooperation of Avalue Technology (Taiwan) was obtained in the provision of the HPC (High Performance Computer) to accelerate and help them produce high-accuracy results.

One requirement for developing and constructing deep learning systems, not just this system, is that it be highly versatile. Therefore, the DataSense system adopted an Intel® processor instead of an NVIDIA GPU. The OS uses Microsoft Windows.

Because the X1 Sequencing application was developed using the OpenVINO™ toolkit, it is expected to be able to seamlessly support future generations of Intel® processors, and enhancements based on fourth-generation Intel® Xeon® processors and the Intel® Arc DG2 (VPU) are being planned.

Compared with the existing analysis process, mostly we offer doctors better-filtered data, easier database comparison, accelerated alignments and variant detection. Biomdcare's solution is capable of reducing cost and 87.5% of time, making this screening become available to the public and whoever need it. Furthermore, it's 800% more efficient and improved accuracy to over 97%.

Results & Impacts

Even though the NGS performance has improved, extremely careful and complicated work is required to handle human genes, and quality control (QC) must be strictly conducted to prevent contamination, and make sure the analysis standard is well-established. The general QC is based on 6 major standards: Read length distribution, GC content, Per base read quality, Nucleotide composition, Per read quality, and Overrepresented sequences.

According to Biomdcare's scope of understanding, confirming the presence of muscular dystrophy gene expression requires roughly two weeks for sequence preprocessing, roughly three weeks for analysis of the massive amount of NGS data, and at least four months (16 weeks) for a decision by experts and doctors, requiring a total of roughly 24 weeks to reach a diagnosis.

Using the system developed by Biomdcare is expected to shorten the previous 24 weeks to about 4 weeks by accelerating the inferencing of the presence of gene expression and by providing highly accurate inference data to experts and doctors. Naturally, the final diagnosis will be performed by a doctor, but the system will likely function as an extremely useful assistant.

We believe that such streamlining is achieved through the high inference accuracy of the low-resource training developed by Biomdcare, the adoption of YOLOv5 to achieve high-speed object detection, the adoption of the OpenVINO™ toolkit to create efficient execution files for Intel processors, the development of the proprietary ZNET framework to integrate YOLOv5 into the OpenVINO™ toolkit, and the adoption of second-generation Intel® Xeon® processors with superior versatility and performance.

In addition to the increasing enhancement of NGS, there is an active movement to try and apply genetic analysis to diagnosis, therapy, and pharmaceutical development, etc. The massive volume of data produced by Whole Exome Sequencing (WES), which reaches as much as 10 GB and even bigger, has become easier to handle and analyze compared to before together with advances in the hardware platform. Moreover, knowledge concerning genetic sequence analysis using deep learning is also being accumulated.

Biomdcare hopes to contribute to medical progress by further enhancing its X1 Sequencing technology and the analysis of hereditary diseases.

The general public believes that single-gene recessive genetic disorders are rare, but objective statistics show that the presence of recessive genetic variants is quite common, even if only ten of the most common single-gene recessive genetic disorders are screened: thalassemia, Juvenile muscular dystrophy, spinal muscular dystrophy, hereditary deafness, haemophilia, phenylketonuria, Wilson's disease, Pompe disease, etc., the individual carrier rates for these disorders range from 1 in 20 to 1 in 100. And the combined carrier rate is as high as 26%, with an average of one in four people carrying the disease. If the number of diseases screened increases to 600, 5 out of 100 couples will have a child with the disease.

In terms of overall incidence, many of the diseases that cause severe malformations or developmental delays occur in children, with about 0.04% of live births having a single gene disorder. Because some diseases do not develop until adulthood, about 8 out of 100 adults have single-gene disorders of varying severity, they are also common in adults. According to the American Genetic Society in 2015, about 8 million children worldwide have severe singlegene disorders.

With the rapid development of bioinformatics, recent years are an exciting period for both genetic technologies and personalised medicine, especially in the area of clinical applications.

The clinical use of next-generation sequencing (NGS) is becoming increasingly mature, initially for non-invasive maternal testing, and is expected to flourish in the field of liquid biopsies and oncology. The clinical application of next-generation sequencing NGS has become more mature. The cancer panel, which is worth mentioning, has been gradually introduced into medical practice in Europe and the United States. Cancer panels will be used for DNA, RNA, Exosome, and even pre-immunotherapy screening.

The advantage is that one test contains multiple biomarkers and multiple indicators, allowing for a more comprehensive analysis of cancer risk. It also has the potential to detect rare and specific variants. However, there are regulatory challenges to overcome including clinical validation, data analysis validation, stability between different types of samples, and variability of results between laboratories, and the need for quality control of each NGS procedure. We believed that more clinical data and standardisation will help to solve these problems.

The importance of pre-conception or prenatal screening

There is no genetic history of recessive genetic disorders. The genes for the disease are hidden in the family for generations, and it is only when both husband and wife happen to be carriers that the genetic disorders of the next generation are caused. Sexually linked and dominant genetic diseases also do not necessarily have a genetic history, as new mutations may arise. Therefore, prenatal screening for single-gene genetic disorders is advisable to avoid the hidden risks over generations.

As long as both husband and wife are carriers of a single gene genetic disorder, the fetus is at risk of developing the rare disease, so pre-conception or prenatal screening for single-gene genetic disorders is necessary. NGS, for example, can test for hundreds of genes or even the entire genome at one time, and is a more comprehensive genetic test that is not limited to specific types of diseases, thus avoiding the risk of passing on defective genes to the next generation.

The future: Assist research community and build standard genomics appliances with easier database comparison

There is currently no cure for Muscular Dystrophy, patients can only have the chance to get involved in clinical trials once they have been diagnosed with certain types. Therefore, it can be helpful for medicine development in the future by providing the AI and DNA database for doctors and researchers.

Biomedica
Simple & Precise

Biomdcare Corpration

Taiwan Headquarter
2F, No.11, Ln. 218, Sec. 2, Zhongxing Rd, Xindian Dist, New Taipei City 231, Taiwan

avalue Technology Inc.

Avalue Technology

Address : 7F, 228, Lian-cheng Road, Zhonghe Dist., New Taipei City 235, Taiwan