Academic Journal of Science and Technology
ISSN: 2771-3032 | Vol. 20, No. 1, 2026

Design of Assisted Cupping Diagnosis Medical System based on FPGA Hardware Acceleration and BP Neural Network

Hantao Sun [1, 3], Long Yang [1, 3], Jin Wang [1, 2], Dehui Zou [1, 2, *], Zhigang Di [1, 3]

[1] North China University of Science and Technology, Tangshan, Hebei, 063210, China
[2] Traditional Chinese Medical College, North China University of Science and Technology, Tangshan, Hebei, 063210, China
[3] College of Electrical Engineering, North China University of Science and Technology, Tangshan, Hebei, 063210, China

*** Corresponding author:** Dehui Zou (Email: 1198916301@qq.com)

Abstract: Cupping diagnosis, a characteristic TCM physical method, serves key roles in preliminary disease screening and constitution identification. Traditional manual cupping suffers from strong subjectivity, descriptive nature and fragmented workflows. To this end, an intelligent cupping diagnosis system based on FPGA hardware acceleration and BP neural network is designed. In this system, FPGA integrates multi-sensor interfaces, and via hardware acceleration completes cupping mark image enhancement, sensor data fusion and synchronous acquisition. A BP neural network diagnostic model is deployed on the upper computer to jointly analyze cupping mark image and internal cup environmental features, achieving intelligent constitution and disease diagnosis. Experimental simulations and theoretical analysis show that the system's temperature measurement error is reduced by ~76% vs. original error, cupping mark image processing hits 79.37 fps, and the BP neural network effectively identifies six constitution types. This research provides technical reference for hardware acceleration and digital upgrading of TCM instruments, and helps advance digital and intelligent development of TCM cupping diagnosis.

Keywords: Medical Canister Diagnosis; FPGA; Hardware Acceleration; BP Neural Network; Intelligent Medical Device.

1. Introduction

The symptoms of diseases in traditional Chinese medicine have strong subjective and descriptive characteristics. In the process of modernization and digitalization of traditional Chinese medicine, objectification and standardization have become the inevitable path for the description of TCM symptoms to move from ambiguity to precision, from complexity to simplicity, and from qualitative to quantitative [1]. Cupping diagnosis, as a distinctive physical diagnostic method of traditional Chinese medicine, achieves the assessment of human health status by analyzing the morphology, color, temperature, humidity, and pressure characteristics of the cupping spots, and bears the important responsibility of initial disease screening and constitution identification [2]. However, with the development of the intelligence and standardization of traditional Chinese medicine, traditional manual cupping diagnosis has gradually exposed shortcomings such as strong subjective dependence, difficulties in quantifying diagnostic indicators, and fragmented diagnostic processes. In this context, intelligent cupping diagnosis assistance devices based on microcontroller units (MCUs) and multiple sensors have emerged, meeting the basic needs for medical data collection and low power consumption.

However, the currently common intelligent cupping diagnosis devices mostly rely on an MCU as the core and have relatively weak parallel processing capabilities, making it difficult to simultaneously solve problems such as the synchronous collection of multi-modal data [3], high real-time image processing and algorithm collaboration, and lacking a design scheme for integrating the "collection - analysis - diagnosis" process into one. Their diagnosis process still relies on the subjective experience of physicians. To address these issues, this study has designed an auxiliary cupping diagnosis medical system based on FPGA hardware acceleration as the data processing foundation and BP neural network as the intelligent diagnosis core, so as to meet the development trend of standardization and intelligence in cupping diagnosis and treatment. This auxiliary cupping diagnosis medical system is supported by the core technologies of FPGA's parallel processing characteristics and BP neural network's intelligent analysis capabilities, balancing real-time performance and intelligence, with controllable development costs and improving the objectivity and automation level of the diagnosis process.

2. System Framework Design

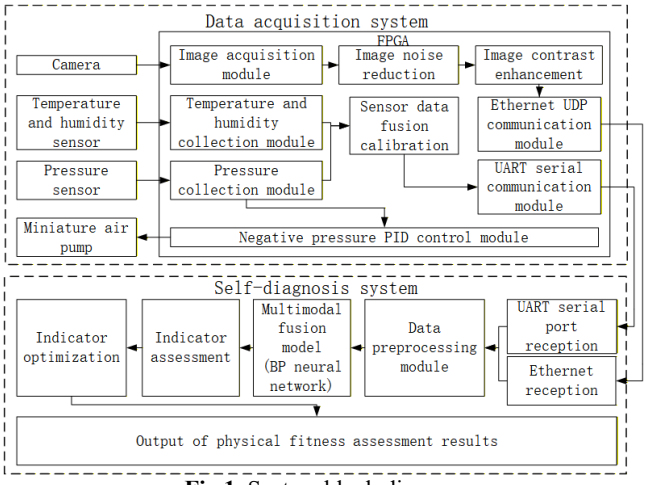

Fig 1. System block diagram

This intelligent tank diagnosis system is composed of two subsystems: data acquisition and autonomous diagnosis. The FPGA serves as the control core for the data acquisition part,

responsible for collecting, processing, transmitting the tank spot images, as well as the temperature, humidity and pressure data inside the tank, and controlling the negative pressure. The upper computer acts as the operating platform for autonomous diagnosis, mainly used to deploy the BP neural network model to conduct diagnosis analysis on the collected feature data. It is responsible for receiving, analyzing and visualizing the feature data. The overall block diagram of the system functions is shown in Fig 1.

3. Principle and Implementation of Data Acquisition System

The data collection from multiple sensors is the data source for the autonomous diagnosis of this system. However, the autonomous diagnosis system has high requirements for the synchronization and accuracy of the input data. Moreover, the sensor chips may be affected by external factors and internal electronic component factors during the data collection and transmission process, resulting in signal noise and data delay. This can easily affect the data collection accuracy and real-time synchronization. Therefore, while designing the basic data collection module, necessary optimization mechanisms should also be added, including synchronous collection, data fusion calibration, image enhancement, and data buffering and metastable state avoidance mechanisms.

Compared with common MCU chips, FPGA can build highly customized sensor interfaces and parallel processing logic circuits independently through a large amount of logic gate resources, memory blocks, and high-performance DSP resources. This not only reduces design redundancy but also significantly improves the real-time data collection and processing performance.

3.1. Multi-sensor Synchronous Acquisition

The multi-sensor data synchronization acquisition part is used to collect the temperature, humidity and pressure data inside the tank. It adopts a modular design to build various communication protocol modules and synchronization trigger modules within the FPGA.

The temperature, humidity and pressure data acquisition uses the DHT11 sensor with single-bus serial transmission and the BMP280 sensor with I2C/SPI standard communication protocol as the acquisition components [4]-[6]. Due to the different communication methods, during the FPGA development process, the standard communication protocols can be customized and modified on the basis of ensuring the basic functions to adapt to the transmission timing of the sensor chips, reducing unnecessary resource occupation during data transmission. The design adopts design ideas such as state machines and pipelines, combined with communication protocol IP cores, to complete the logically rigorous single-bus and I2C bus designs.

The synchronization trigger module adopts a synchronous timing design, mainly used to generate the trigger signals required for each sensor acquisition. The FPGA uses the phase-locked loop (PLL) to multiply the 50 MHz clock signal input from the external crystal oscillator to 100 MHz as the system reference clock. The standard controllable trigger signal generated by the FPGA based on the second pulse is used as the standard trigger signal for the sensors. The sensor trigger calibration logic uses this standard trigger signal as the reference signal combined with the sensor's own working frequency to generate a precise trigger signal for triggering the sensor to work, minimizing the impact of sensor trigger delay on the subsequent data fusion and the effectiveness of neural network model analysis. At the same time, it is conducive to obtaining the timestamp when each sensor is triggered, facilitating the alignment of timestamps before data transmission[7].

3.2. Sensor Data Fusion Calibration

The processing of sensor data mainly improves the accuracy of a single data through data compensation and heterogeneous data fusion, minimizing the random error of the original signal.

In view of the low measurement accuracy and large fluctuations of DHT11 temperature, this design scheme linearly compensates the original DHT11 temperature data TDHT11 based on the zero bias temperature coefficient ΔT obtained during the static calibration of DHT11, as shown in Equation (1). At the same time, since both BMP280 and DHT11 can output temperature data, a weighted fusion algorithm as shown in Equation (2) can be adopted to fuse the data of the two sources respectively with weights of ω1 and ω2 to reduce the random error of the original DHT11 data.

$$T_{DHT11\_cal} = T_{DHT11} - \Delta T \tag{1}$$

$$T_f = \omega_1 \times T_{BMP280} + \omega_2 \times T_{DHT11\_cal} \tag{2}$$

3.3. Image Enhancement Algorithm

The camera uses the OV5640 CMOS sensor produced by OmniVision Company. This sensor uses SCCB and DVP interfaces for sensor configuration and image data transmission. As is well known, the SCCB bus is very similar to the I2C bus. The main difference is that the SCCB has removed the continuous read-write function of I2C [8]. Therefore, the FPGA can directly customize the I2C communication bus to meet the requirements of the SCCB bus.

However, the raw images collected by the CMOS sensor inevitably have noise interference, edge blurring, and format redundancy due to environmental light or factors of the chip itself. Thus, an image enhancement module is designed to use digital logic resources and pipeline design to achieve the hardware-level implementation of the median filtering algorithm and histogram equalization algorithm.

Median filtering is a nonlinear smoothing technique. It sets the single-channel value or gray level of each pixel point to the median of all pixel points in the neighborhood window of that point[9]. Median filtering has excellent suppression effect on salt-and-pepper noise in the image and can largely retain the detailed features of the image.

For RGB color image median filtering in FPGA implementation, the idea is to separate the RGB channels and perform median filtering on each channel separately. Compared with the vector median filter (VMF) algorithm commonly used for high-quality color noise reduction[10], this implementation scheme can avoid the resource overhead caused by large-scale vector operations while ensuring image quality. The specific implementation of the median filtering

algorithm on the FPGA platform is shown in Fig 2. First, the RGB three-channel data is separated, and then the pixels of the current row are concatenated with the pixels of the row to build a 3×3 sliding window for each single channel independently. On this basis, the pairwise comparison method (also known as "tournament sorting") is adopted to complete the median solution for each color channel using a three-level comparator. Finally, the filtered results of the three channels are recombined into an RGB image for output.

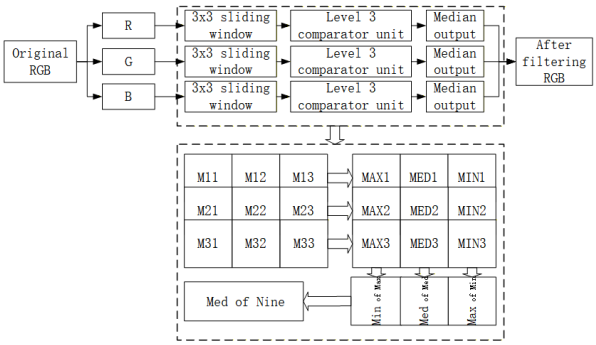

Fig 2. Principle diagram of median filtering algorithm for channel separation and recombination

The histogram equalization algorithm is actually a gray-scale transformation process. It uses the information of the image for transformation and transforms the current gray-scale distribution through the transformation function into a wider and more uniform image, expanding the dynamic range of the image to enhance the image contrast. However, since the RGB color space is an additive color mixing model and the three channels are strongly coupled, directly performing equalization on R/G/B will destroy the color ratio between the channels. Therefore, for the processing of the image on the FPGA platform, RGB can be first converted to YCbCr, and the Y channel can be separately equalized before returning and calculated as an RGB color space image. In color conversion, a 256 times amplification method is used, as shown in Equation (3), to quantize the floating-point operations into integer operations for solving the values of YCbCr channels. Then, Block RAM is used as the histogram cache, and the pixel gray levels are re-mapped using Equation (4) to stretch the brightness dynamic range [11]. Finally, the Y channel is re-fused into YCbCr and inversely converted back to an RGB image for output, achieving contrast enhancement of the color image and ensuring that the skin texture features are not blurred due to changes in light.

$$\begin{cases} Y=(77R+150G+29B)>>8 \\ Cb=(128B-85G-43R+32768)>>8 \\ Cr=(128R-107G-21B+32768)>>8 \end{cases} \tag{3}$$

$$D_B = f(D_A) = \frac{D_{\max}}{A_0} \cdot \sum_{t=0}^{D_A} H(i) \tag{4}$$

3.4. Data Buffering and Metastable State Avoidance

To address the issue of inconsistent data output rates from sensors, an advanced first-in-first-out (FIFO) buffer mechanism based on "domain partitioning cache, timestamp alignment, and synchronous read enable" was designed on the FPGA. Synchronous FIFOs and Block RAM storage blocks were configured separately for DHT11 and BMP280 to cache the environmental data inside the tank. The synchronous trigger signal inside the FPGA was bound with the timestamp and written to the first addresses of both FIFOs. Through a hardware state machine, the completion status of data acquisition was determined. After both data streams were fully collected, the FIFO read operation was enabled synchronously to achieve synchronous output of multi-sensor data.

The working clocks for image acquisition and Ethernet transmission are often independent. Direct transmission is prone to introducing metastable states and causing data errors and packet loss. To overcome this, the asynchronous FIFO, which has independent read and write clocks and pointer synchronization logic (such as Gray code encoding), is utilized for cross-clock domain transmission, avoiding metastable states[12]. The image data is written to the FIFO according to the acquisition clock, while the Ethernet module reads the data according to its own clock, achieving stable and reliable transmission of data between different clock domains.

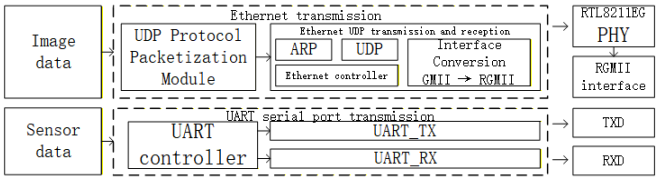

Fig 3. Data transmission link architecture

3.5. Data Transmission Interface Design

The system adopts a dual-link hybrid architecture of Ethernet and serial port as shown in Fig 3, which respectively meets the bandwidth and rate requirements for high-speed image data transmission and low-speed sensor data transmission. The gigabit Ethernet link is centered around the RTL8211EG Ethernet PHY chip, and is implemented by FPGA through hardware logic to encapsulate the lightweight UDP protocol. After performing the ping-pong caching operation on the pre-processed image data, it forms data frames in the format of "frame header + image resolution + pixel data + CRC32 check", and completes high-speed and reliable transmission via the RGMII interface [14]. The serial port link is responsible for the interaction between sensor data and the upper computer control instructions. The FPGA uses the UART IP core and adopts the "frame header + quantized data + checksum" packaging structure to establish stable two-way communication with the upper computer.

4. Principle and Implementation of Autonomous Diagnosis System

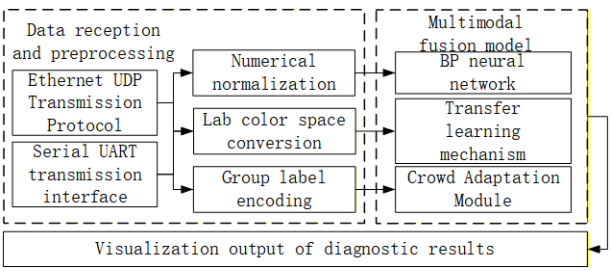

Fig 4. Autonomous diagnostic system block diagram

The relationship between disease diagnosis and medical data is mostly nonlinear, high-dimensional and fuzzy, and general processing algorithms cannot achieve this goal. However, the BP neural network can fit complex mapping

relationships and break through the limitations of traditional linear models. As shown in Fig 4, the autonomous diagnosis system takes the multimodal fusion model as the core, combined with the data reception preprocessing module and the diagnosis result visualization output module to form a complete upper computer diagnosis system.

4.1. Data Reception and Preprocessing

The host computer of the autonomous diagnosis system also adopts a dual-link data reception structure. Based on the Socket interface and the PySerial library, it respectively acquires the tank spot images and environmental sensor data. The host computer receives the pre-processed tank spot image data transmitted from the FPGA end through the UDP protocol, decodes it, and restores it as a 640×480 resolution RGB image and completes the Lab color space conversion, providing core visual feature support for disease diagnosis; it receives the fused and calibrated environmental parameters through the serial port, decodes them, and outputs standardized parameter values through normalization processing, providing environmental correlation basis for diagnosis analysis. To ensure the reliability of data transmission, CRC32 checksum and data frame accumulation and verification are set for image and sensor data transmission respectively, invalid data is discarded, and the integrity and accuracy of the input data to the diagnosis model are ensured.

4.2. Multimodal Fusion Model

The multimodal diagnostic model of this autonomous diagnostic system is centered on the BP neural network, combined with the transfer learning mechanism and the population adaptive module, to achieve the joint diagnosis of spot images and environmental sensor data.

The input layer of the model receives two types of modal data. One is the spot image processed by image enhancement and Lab color space conversion, which extracts texture and color by separating the luminance component L and the chromatic components a and b, and then normalizes it to form a 12-dimensional image feature vector XI. The other is a 6-dimensional environmental feature vector XE composed of weighted fused environmental parameters and corresponding change rates. The image features and environmental features are concatenated to form a 18-dimensional fused feature vector [XI, XE], which is used as the input of the BP neural network.

The BP neural network structure used in this paper is shown in Fig 5, consisting of an input layer, a hidden layer, and an output layer. The input layer contains 18 neurons corresponding to each fused feature dimension; the hidden layer consists of a single layer of 25 neurons, using the ReLU activation function; the output layer is composed of 6 types of physical constitution labels: damp heat, blood stasis, qi deficiency, cold dampness, qi-blood deficiency, and normal.[13]-[15]

To enhance the generalization ability of the model in small sample clinical scenarios, the transfer learning mechanism is introduced in the design. The network weights pre-trained on the public dermatological dataset are used as the initialization parameters, and the model is adapted to the spot diagnosis task through a fine-tuning strategy to alleviate the overfitting problem caused by the scarcity of clinical samples. At the same time, for the differences in spot characteristics among different populations, the input features are encoded with population labels and dynamically weighted to optimize the diagnostic sensitivity and specificity of the model for different subgroups.

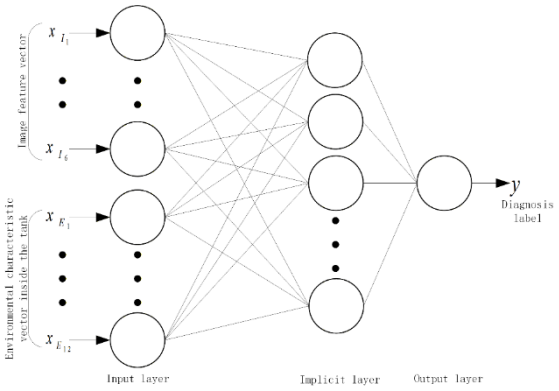

Fig 5. BP neural network structure diagram

4.3. Visualization Output of Diagnostic Results

At the output end of the model, the confidence level of the diagnosis results is calibrated with a threshold. If the result is below the threshold, a secondary review mechanism is triggered to enhance the reliability and clinical acceptability of the diagnosis results, as well as the accuracy and robustness of the autonomous diagnosis. Subsequently, a graphical interface developed based on the React framework presents the diagnosis result with the highest confidence level and the curve of the internal tank environmental parameters. The system uses the MySQL database to store user information and historical diagnosis data, and also has the capabilities of data query and long-term storage.

5. Experimental Simulation and Theoretical Analysis

After optimizing the data weighted fusion algorithm in Section 2.2, the random error of the temperature in a single dimension was reduced. According to the official data of the sensors, the typical values of the original temperature measurement errors of BMP280 and DHT11 are σBMP280 = ±0.5℃ and σDHT11 = ±2℃, respectively. Based on the theoretical calculation of obtaining the fusion weights ω1 = 0.9412 and ω2 = 0.0588, and by using the error propagation law formula (5), the theoretical error of the weighted fusion temperature can be calculated to be approximately 0.485℃. Compared with the original measurement error of DHT11, it is reduced by about 76%, which is obviously beneficial for the optimization of the temperature data acquisition accuracy of DHT11.

$$\sigma_f = \sqrt{\omega_1^2 \times \sigma_{BMP280}^2 + \omega_2^2 \times \sigma_{DHT11}^2} \tag{5}$$

In terms of image enhancement, compared with the traditional MCU serial processing scheme, the parallel architecture and pipeline design based on FPGA can significantly reduce the image processing delay. For the median filtering algorithm, for a single gray-scale pixel, using serial bubble sort processing requires at least 8 to 10 clock cycles, and it requires CPU computing resources and is managed by the task scheduling system, resulting in relatively high processing delay. However, this design adopts a pipeline and multi-channel parallel processing method. On average, it only takes 1 clock cycle to complete the median extraction of a single color channel and a single pixel, and it is not affected by the task scheduling, allowing for full-speed processing throughout. To more intuitively evaluate the image processing

efficiency, a simulation stimulus file written in Verilog HDL language is used to simulate the video transmission timing of the camera and load the test video. The image processing module is fully simulated using the simulator provided by Vivado. The simulation results are shown in Fig 6. When inputting an image with a resolution of 640×480, from the input of the first frame image to the effective output of the first frame image, the module takes approximately 12.6ms, and the processing frame rate is about 79.37fps. Most of the processing delay is generated in the inherent delay of the pipeline. The processing frame rate of nearly 80fps fully meets the real-time requirements of the image acquisition and processing of this system.

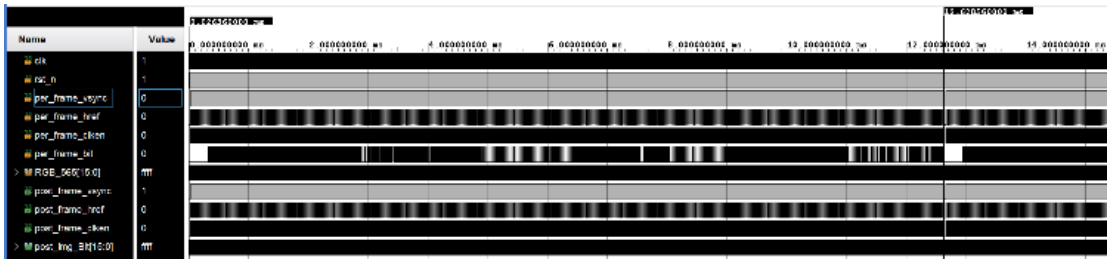

Fig 6. Image processing module simulation results

This intelligent cupping diagnosis system takes advantage of the hardware-level parallel and high-speed data processing capabilities of FPGA devices to complete the collection and enhancement processing of cupping spot images as well as temperature, humidity, and pressure data. This provides a stable data stream with high reliability and strong real-time performance for the subsequent autonomous diagnosis system. The BP neural network effectively fits the complex nonlinear disease feature relationships in traditional Chinese medicine diagnosis, achieving objective diagnosis and treatment course analysis to a certain extent.

Fig 7. System operation results

The overall trial operation results of the intelligent cupping diagnosis system are shown in Fig 7. The system interface presents historical data within the patient's treatment course, combined with current data analysis, to determine the patient's constitution type, risk level, and corresponding recommended treatment cycle. The detailed historical data and constitution diagnosis results can provide data references for professional physicians to formulate personalized treatment plans for patients.

6. Conclusion

This paper addresses the core issues of traditional cupping diagnosis, which are highly subjective, insufficient quantification, and fragmented processes. It proposes and implements an intelligent cupping diagnosis system based on FPGA hardware acceleration and BP neural network, fully integrating the parallel processing and customization features of FPGA with the flexibility of the upper computer deploying the BP neural network algorithm. The customized development for various sensor interfaces such as temperature, humidity, pressure, and images was completed on the FPGA, as well as the optimization logic design including data calibration and synchronous acquisition. On the basis of high-definition image acquisition, the hardware-level acceleration of image median filtering and histogram equalization algorithms was achieved on the FPGA platform, achieving the effect of suppressing noise and improving image contrast. Compared with the traditional serial processing scheme, the processing efficiency was significantly improved. The BP neural network deployed on the upper computer serves as the core of diagnosis and combines the transfer learning method to achieve objective assessment. This research not only breaks through the limitation of traditional cupping diagnosis relying too much on the subjective experience of physicians, but also provides reusable technical references for the intelligent and digital development of traditional Chinese medicine cupping diagnosis. In the future, this research will focus on data set expansion, algorithm optimization, and hardware convenience upgrades, continuously improving the universality and clinical adaptability of the system, and promoting the acceleration of traditional Chinese medicine cupping diagnosis towards digitalization and large-scale development.

Acknowledgments

Hebei Province Project for Research and Practice on Higher Education Teaching Reform (2025GJJG230); Hebei

Province Project for College Students' Innovation and Entrepreneurship Training (S202510081100).

References

[1] Fu Haoran, Dai Guohua, Gao Wulin, et al. Approaches and Methods for Selecting Representative Diseases in Digital Research on Enhancing Syndrome Differentiation and Treatment in Traditional Chinese Medicine [J]. World Science and Technology – Modernization of Traditional Chinese Medicine, 2021, 23(04):1205-1210.

[2] Qin Xian, Shen Pengfei. The Clinical Significance of 'Pot Spots' [J]. Shanghai Journal of Traditional Chinese Medicine, 2011, 45(8):20-21.

[3] Khan I M ,Silva D B .Harnessing FPGA Technology for Energy-Efficient Wearable Medical Devices[J]. Electronics, 2024, 13(20):4094-4094. DOI:10. 3390/ ELECTR ONICS132 04094.

[4] Liang Taohua, Zhou Jiang, Xiao Guannan. Design and Implementation of a Thermohygrometer Based on DHT11 [J]. Electronic Production, 2024, 32(15):88-90.

[5] Niu Wenju, Huang Rongyu, Han Jianqiang. FPGA-Based Micro-Sensor Signal Acquisition System [J]. Sensors and Microsystems, 2019, 38(05):104-106.

[6] Miftahul K, Harditio A P, Pande W , et al.Design of Air Pressure Measuring Devices Using a Barometric Pressure 280 (BMP280) Sensor Based on Arduino Uno[J].BULETIN FISIKA,2020,21(1):14-14.DOI:10.24843/bf.2020.v21.i01.p03.

[7] Huang Zhanhua, Li Xiaowei, Wang Kangnian, et al. Design of a Multi-Sensor Fusion Synchronous Timing System Based on FPGA [J]. Instrument Technology and Sensors, 2023, (7):52-61.

[8] Yuan Weiqi, Tang Yonghua. Implementation of OmniVision Image Sensor SCCB Bus Protocol in DSP [J]. Journal of Instruments and Instrumentation, 2006, (S2): 1687-1688.

[9] Zhang Conghua, Yang Yong, Liu Ruofan, et al. Application of Graphic and Image Technology in Infrared Thermal Imaging [J]. China Testing, 2012, 38(2):77-80.

[10] Deng Tingquan, Dong Tianzhen, Xie Wei, et al. Adaptive center-weighted median filtering method for color images [J]. Control and Decision, 2013, 28(9):1372-1376.

[11] Huang Xiaoxin, Zhu Zhengtai, Li Jingtao. Underwater Image Enhancement Algorithm Based on Weighted Fusion and FPGA Design [J]. Electronics Technology, 2025, 54(1):106-110.

[12] Jiang Junlun, Feng Daqiang, Xu Xinrui, et al. Design of an Image Acquisition and Processing System Based on FPGA GPU [J]. Computer Measurement and Control, 2023, 31(8): 273-279, 305.

[13] Fu Hao, Sun Heng, Zhao Zhongkai. Measurement of Radar Signal Carrier Frequency Based on BP Neural Network [J]. Ship Electronic Warfare, 2024, 47(2):71-75.

[14] Hong Zhiwei. Design of Gigabit Ethernet Image Acquisition and Processing System Based on FPGA [D]. Jishou University, 2022.

[15] Zhao Yuan. Research on Digital Image Classification System Based on BP Neural Network [J]. Information Technology and Informatization, 2023, (12): 123-127.