UDC 681.5:613

Li Jinqiong

MATHEMATICAL MODELING OF DIAGNOSTICS OF ACUTE FORMS OF LEUKEMIA BASED ON MORPHOLOGICAL DATA

Department of Biomedical Engineering and Optic-Electronic Systems,
Vinnytsia National Technical University
Department of Hematology and Oncology, Lijiang People's Hospital, Yunnan, China

Abstract. *From the scientific novelty of the work, a complex combination of statistical and heuristic models is proposed, morphological criteria of FAB-classification are formalized in the form of mathematical dependencies, a universal approach to automated diagnostics is developed. The practical value lies in the fact that the theoretical results allowed to increase the accuracy of diagnostics, reduce the subjectivity of medical decisions, the possibility of integration into medical information systems, and use in decision support systems.*
Keywords:acute leukemia, AML, mathematical modeling, classification, logistic regression, Bayesian approach

Анотація. *Запропоновано комплексне поєднання статистичних та евристичних моделей, формалізовано морфологічні критерії FAB-класифікації у вигляді математичних залежностей, розроблено універсальний підхід до автоматизованої діагностики. Практична цінність полягає в тому, що теоретичні результати дозволили підвищити точність діагностики, зменшити суб'єктивність лікарських рішень, можливість інтеграції в медичні інформаційні системи, здійснювати використання в системах підтримки прийняття рішень.*
Ключові слова: гострий лейкоз, AML, математичне моделювання, класифікація, логістична регресія, байєсівський підхід

INTRODUCTION

Acute leukemias are a group of severe oncohematological diseases characterized by rapid progression and high mortality without timely diagnosis. FAB classification (M0–M7) is based on morphological features of cells and the proportion of blasts, however, the traditional assessment is subjective and depends on the experience of the doctor. This necessitates the use of mathematical models to increase the accuracy of diagnosis [1, 2, 3].

Purpose of the work

Development of mathematical models for the classification of acute forms of leukemia based on the morphological characteristics of bone marrow cells.

1. MATHEMATICAL MODELS FOR THE DIAGNOSIS OF ACUTE LEUKEMIA

The study is based on morphological data of AML subtypes (M0–M7), in particular:

- proportion of blasts;
- presence of Auer bodies;
- degree of cell differentiation;
- dominant cell line.

The following were used for modeling:

- multiclass logistic regression;
- Bayesian classification;
- threshold rules;
- k-means method.

1.1. Features (input parameters)

The model uses the following features: % blasts; MPO (myeloperoxidase); Auer bodies; granulation; cell lines (monocytes, erythroids, megakaryocytes) (Table 1) [4, 5, 6].

Table 1. Morphologic Features and Blast Proportions of AML FAB Subtypes

FAB subtype	**Major cell proportion criteria (bone marrow)**	**Blast proportion**	**Key morphologic features (high power microscopy)**
AML-M0	≥30% blasts; MPO negative or very low	≥30%	Very primitive blasts; high nuclear-cytoplasmic ratio; scant cytoplasm; almost no granules; Auer rods usually absent; resembles lymphoblasts
AML-M1	≥90% of non-erythroid cells are myeloblasts	≥30%	Medium–large blasts; prominent nucleoli; fine chromatin; few azurophilic granules; Auer rods may be present
AML-M2	30–89% blasts with granulocytic maturation	≥30%	Myeloblasts with promyelocytes and myelocytes; Auer rods common; cytoplasmic granules more evident
AML-M3	Predominantly abnormal promyelocytes	Variable	Cells with abundant coarse granules; bundles of Auer rods (faggot cells); nuclei often bilobed or kidney-shaped
AML-M4	Granulocytic and monocytic components each ≥20%	≥30%	Mixed myeloblasts and monoblasts; monocytic cells have abundant cytoplasm; folded nuclei; M4Eo shows abnormal eosinophils
AML-M5	≥80% monocytic lineage cells	≥30%	Large monoblasts; abundant basophilic cytoplasm; frequent vacuoles; irregular or folded nuclei
AML-M6	≥50% erythroid precursors; ≥30% blasts among non-erythroid cells	≥30% (non-erythroid)	Marked increase of proerythroblasts and early erythroblasts; deeply basophilic cytoplasm; cytoplasmic vacuoles; prominent erythroid dysplasia
AML-M7	≥50% blasts are megakaryoblasts	≥30%	Basophilic cytoplasm; characteristic cytoplasmic blebs or budding; marrow fibrosis common; abnormal platelets

Based on the table, we form a vector of patient characteristics [5, 7]:

$$X = (x_1, x_2, \ldots, x_n),$$

where:

- x1 — proportion of blasts (%);
- x2 — MPO activity (0/1 or scale);
- x3 — presence of Auer bodies (0/1);
- x4 — degree of granulation;
- x5 — morphological type of cells;
- x6 — proportion of monocytic cells;
- x7 — proportion of erythroid cells;
- x8 — signs of megakaryoblasts.

Probability of belonging to subtype *k:*

$$P(Y = k|X) = \frac{P(X|Y = k)P(Y = k)}{\sum_j P(X|Y = j)P(Y = j)}$$

The model allows taking into account a priori subtype frequencies and working with incomplete data

1.2. Threshold diagnostic model

A threshold diagnostic model is proposed for a quick clinical decision

$$\boldsymbol{Y} = \begin{cases} \boldsymbol{M_o}, & if; \\ \boldsymbol{M_3}, & if\ ; \\ \boldsymbol{M_5}, & if; \\ \boldsymbol{M_6}, & if; \\ \boldsymbol{M_7}, & if. \end{cases}$$

In this case, the probability of belonging to subtype *k:*

$$P(Y = k|X) = \frac{e^{\beta_k^T X}}{\sum_j e^{\beta_j^T X}}$$

Used for: automatic diagnosis of FAB subtypes (M0–M7); clinical decision support

1.3.Clustering model

$$min \sum_{i=1}^{N} \sum_{k=1}^{K} w_{ik} \|X_i - \mu_k\|^2$$

Used for: Identifying new subtypes; • Refining classification

2. EXPERIMANTAL RESULTS

A mathematical classification model was constructed that allows determining the subtype of leukemia based on the vector of patient characteristics.

The logistic model showed high efficiency in distinguishing subtypes:

- M_0–M_2 — by the level of differentiation;
- M_3 — by the presence of Auer bodies;
- M_4–M_5 — by the monocytic component;
- M_6–M_7 — by a specific cell line.

The Bayesian approach provided flexibility in the case of incomplete data.

The threshold model allows for rapid clinical interpretation without complex calculations using such an algorithm.

```
# Signs:
# Tags:
# blasts (%) – blast percentage
# MPO – activity (0/1)
# auer_rods – presence (0/1)
# granularity – granulation level (0-2)
# monocytic – monocytic percentage (%)
# erythroid – erythroid percentage (%)
# megakaryoblast – presence (0/1)
data = [
# blasts, MPO, auer, granularity, monocytic, erythroid, megakaryoblast, label
[35, 0, 0, 0, 5, 5, 0, "M0"],
[85, 1, 1, 1, 5, 5, 0, "M1"],
[60, 1, 1, 2, 10, 5, 0, "M2"],
[40, 1, 1, 2, 5, 5, 0, "M3"],
[50, 1, 0, 1, 30, 5, 0, "M4"],
[45, 1, 0, 1, 85, 5, 0, "M5"],
[40, 1, 0, 0, 5, 60, 0, "M6"],
[35, 1, 0, 0, 5, 5, 1, "M7"],
columns = [
"blasts", "MPO", "auer_rods", "granularity",
"monocytic", "erythroid", "megakaryoblast", "label"
df = pd.DataFrame(data, columns=columns)
# ---------------------------
# 2. Data preparation
# ---------------------------
X = df.drop("label", axis=1)
y = df["label"]
# separation
X_train, X_test, y_train, y_test = train_test_split(
x, y, test_size=0.25, random_state=42
# --------------------------
# 3. Model building
# --------------------------
model = RandomForestClassifier(
n_estimators=100,
max_depth=5,
random_state=42
model.fit(X_train, y_train)
```

3. CONCLUSIONS

From the scientific novelty of the work, a complex combination of statistical and heuristic models was proposed, morphological criteria of FAB-classification were formalized in the form of mathematical dependencies, and a universal approach to automated diagnostics was developed.

The practical value lies in the fact that the theoretical results allowed to increase the accuracy of diagnostics, reduce the subjectivity of medical decisions, the possibility of integration into medical information systems, and use in decision support systems.

Morphological data can be effectively formalized in the form of mathematical models. The most accurate are multi-class statistical models. The combination of different approaches provides an effective result of assessing the levels of pathologies of acute forms of leukemia based on morphological data. The proposed models can be used to automate the diagnosis of acute leukemia.

REFERENCES

1. Oleg Avrunin, Yurii Samokhin, Konrad Gromaszek, Vovodymir Dudenko, and Li Jinqiong, etc. "Optimization of cell annotation process: combining manual and automatic labeling in biomedical data analysis optoelectronic systems", Proc. SPIE 14009, Photonics Applications in Astronomy, Communications, Industry, and High Energy Physics Experiments 2025, 140090L (30 December 2025); https://doi.org/10.1117/12.3096280
2. Li Jinqiong, S. Pavlov, O. Poplavskyi,O. Kaduk. Construction Guidelines for Optical-Electronic Expert Systems in Blood Rheology, ІТКІ, вип. 60, вип. 2, с. 107–121, Жов 2024.
3. Stoltz, J.F., Streiff, F., Larcan A. et al. Interetde la mesure de la deformabilite et del'agregation des hematies en biologie clinique. Ann. med. Nancy et Est. 26(6), 431—436.
4. Timchenko,L.I. Pavlov,S.V., Wójcik,W., et al.(2017).Bio-inspired approach to multistage image processing, Proc. SPIE 10445,Photonics Applications in Astronomy, Communications, Industry, and High Energy, 104453M.Wong, T.Y., Klein, R., Sharrett A.R., et al.(2014).Retinal arteriolar diameter and risk for hypertension.Ann. Intern. Med, l40(4), 248-255.
5. Ushenko, A., Dubolazov, A., Zheng, J. et al. “3D polarization-interference holographic histology for wavelet-based differentiation of the polycrystalline component of biological tissues with different necrotic states. Forensic applications,” J. Biomed. Opt. 29(5), 052920 (2024).
6. Wójcik, W., Pavlov, S., Kalimoldayev, M., “Information Technology in Medical Diagnostics II,” London: Taylor & Francis Group, CRC Press, Balkema book, p. 336 (2019).
7. Ushenko, A.G., Zhytaryuk V.G., Vanchulyak, Ya. O. et al., "Statistical and cross-correlation structure of Jones-matrix images of polycrystalline films of biological fluids," Proc. SPIE 10977, Advanced Topics in Optoelectronics, Microelectronics, and Nanotechnologies IX, 109773T, December 2018. https://doi.org/10.1117/12.2323586.

ЛІ ЦЗІНЬЦЮН - аспірант кафедри біомедичної інженерії та оптико-електронних систем,Вінницький національний технічний університет, м. Вінниця, Хмельницьке шосе, 95, Україна, Відділення гематології та онкології, Ліцзянська народна лікарня, Юньнань, Китай, e-mail: 173844678@qq.com, ORCID:0000-0002-0505-621X

Науковий керівник: ПАВЛОВ СЕРГІЙ – Заслужений діяч науки і техніки України, д.т.н., професор, професор кафедри біомедичної інженерії та оптико-електронних систем, науковий керівник ННЛ біомедичної оптики та фотоніки, вінницький національний технічний університет, *e-mail: psv@vntu.edu.ua*

LI JINQIONG - Postgraduated Student of Biomedical Engineering and Optic-Electronic Systems Department, Vinnytsia National Technical University, 95, Khmelnitske shosse, 21021, Vinnytsia, Ukraine, Department of Hematology and Oncology, Lijiang People's Hospital, Yunnan, China, e-mail: *173844678@qq.com*, ORCID:0000-0002-0505-621X

Scientific supervisor - PAVLOV SERGII – Honored Worker of Science and Technology of Ukraine, Doctor of Technical Sciences, Professor, Professor of the Department of Biomedical Engineering and Optoelectronic Systems, Scientific Director of the National Laboratory of Biomedical Optics and Photonics, Vinnytsia National Technical University, *e-mail: psv@vntu.edu.ua*