**Title**: *Systems and Methods for Cancer Classification Using Multi-Modal Transformer-Based Architectures*

**Inventor**: Dr. R. Craig Stillwell

**Filing Date**: July 20, 2025

**1. FIELD OF THE INVENTION**

This invention relates to the field of biomedical informatics and machine learning. Specifically, it describes a system and method for the classification of cancer types using multi-modal genomic data integrated through transformer-based artificial intelligence architectures.

**2. BACKGROUND OF THE INVENTION**

Accurate classification of cancer types is essential for diagnosis, treatment planning, and personalized medicine. Although multiple high-throughput biological datasets are now available — including transcriptomics (e.g., RNA-seq), DNA methylation, and histone modification signals — few AI systems leverage these diverse modalities in a unified, explainable, and production-ready way.

Traditional diagnostic models either use unimodal data or are research prototypes that lack deployment infrastructure, scalability, or clinical interpretability. This invention addresses that gap.

**3. SUMMARY OF THE INVENTION**

The invention, referred to as **Cancer Alpha**, is a production-ready, multi-modal AI system for cancer classification. It uses multiple sources of biological data, processed through domain-specific encoders, fused using a transformer-based architecture, and passed through an ensemble classifier for high-accuracy predictions. The system includes an interpretability layer to produce feature-level importance rankings and a fully containerized infrastructure for deployment in clinical or research settings.

Cancer Alpha is deployed using FastAPI, containerized via Docker and Kubernetes, monitored using Prometheus and Grafana, and integrated with secure APIs for clinical usage.

**4. DETAILED DESCRIPTION OF THE SYSTEM**

**4.1 Data Sources**

Cancer Alpha ingests and harmonizes data from:

* Gene expression datasets (e.g., TCGA RNA-seq)
* DNA methylation data (e.g., Illumina arrays)
* Histone modification profiles (e.g., ENCODE ChIP-seq)

**4.2 Preprocessing Pipeline**

* Each modality is independently normalized, scaled, and filtered.
* Missing values are imputed using modality-specific techniques.
* Dimensionality reduction is applied (e.g., PCA, feature selection).

**4.3 Architecture**

* **Modality-specific encoders** extract latent feature representations for each data type.
* A **multi-modal transformer fusion layer** performs cross-modal learning using self- and cross-attention.
* The fused embeddings are passed to an **ensemble classification layer** (e.g., Random Forest, XGBoost) that outputs the predicted cancer type.
* An **attention-based interpretability layer** ranks the biological features contributing to the decision.

**4.4 Output**

* Predicted cancer classification (e.g., lung, breast, colon)
* Ranked list of contributing features (genes, methylation sites, histone markers)

**4.5 Deployment Infrastructure**

* FastAPI-based backend for real-time inference
* REST API endpoints for integration with lab systems or diagnostic interfaces
* Containerization via Docker; orchestration via Kubernetes
* Prometheus and Grafana for real-time monitoring and system observability
* Redis cache layer for inference speed and efficiency
* CI/CD pipeline implemented via GitHub Actions for reproducible deployment

**5. USE CASES**

* Early diagnostic support for oncologists using minimally invasive samples
* Multi-omics research platforms exploring biomarkers and pathway analysis
* Integration into lab-based or hospital-based clinical workflows
* Cloud-based cancer inference services for precision oncology

**6. LEGAL NOTES**

This application is a **provisional patent application** and does not include formal claims. However, the invention disclosed herein is intended to establish intellectual property priority for:

* The integration of multi-modal omics via a transformer-based AI architecture
* The inclusion of an explainability layer for feature attribution
* The containerized deployment of said model using real-time APIs and monitoring tools