**Discusión**

Según los resultados obtenidos, se puede analizar que en la información revisada no existe una metodología en ciencia de datos definida para el diagnóstico del cáncer de mama. En efecto, La mayoría de literatura científica apunta directamente al uso de técnicas de ML y DL para el diagnostico o pronostico del cáncer de mama exponiendo el nivel de precisión, cantidad de falsos positivos, gasto computacional y modelos algorítmicos utilizados para determinar el posible padecimiento de esta enfermedad. Y aunque estas investigaciones, brindan información valiosa para mejorar la precisión, sensibilidad y especificidad de las técnicas de ML y DL en el diagnóstico del cáncer, carecen de una metodología clara en donde la idea principal gire entorno de la comprensión del dominio y la toma de decisiones por parte de los oncólogos. En particular, la mayoría de investigaciones llegan a resultados en términos de precisión y exactitud, pero no se profundiza en el valor real que el medico especialista le pudo dar a los datos para tomar una decisión y el impacto que dicha decisión tiene en la usabilidad del modelo generado, a sabiendas que el experto es quien finalmente evalúa si los resultados obtenidos por los algoritmos son veraces y permiten diagnosticar de forma ágil el cáncer de mama generando un valor agregado que cumpla con los objetivos del dominio del negocio. Por consiguiente, aunque la comunidad de investigación en ciencia de datos este en crecimiento constante, esté explorando nuevos dominios, creando nuevos roles especializados y este realizando un gran esfuerzo de investigación para desarrollar análisis avanzados, mejorar modelos de datos y generar nuevos algoritmos apoyados de los campos de las matemáticas, la estadística y la informática, estas habilidades no son suficientes para la aplicación de la ciencia de datos en proyectos reales (Martinez2021) , puesto que la mayoría de proyectos basados en datos presentan problemas organizativos y socio-técnicos que surgen al ejecutar un proyecto de ciencia de datos: falta de visión y objetivos claros, un énfasis sesgado en cuestiones técnicas, ambigüedad de los roles. Dicho lo anterior, aunque las investigaciones seleccionadas en la revisión sistemática no giren en torno al cáncer de mama, proponen fases que abarcan aspectos relevantes en la organización, planteamiento y ejecución de un proyecto en ciencia de datos que permiten generar valor significativo a los datos teniendo como eje el dominio del negocio. Cabe resaltar que la metodología CRIPS-DM se selecciono debido a que comprende todas las fases básicas que debe cumplir cualquier proyecto que tenga en su contexto el análisis de datos, y aunque sea la metodología más utilizada, carece de lineamientos que profundicen en la organización de los equipos de trabajo para llevar a cabo procesos de gestión que se alienen con el software y las metodologías de desarrollo agiles. En conclusión, tras el esfuerzo por integrar los resultados analizados en este trabajo, parece bastante plausible afirmar que no existe una metodología puntual en ciencia de datos que se enfoque el diagnostico del cáncer de mama, más sin embargo, diversos autores proponen metodologías que permiten la aplicación de la ciencia de datos en proyectos reales y que pueden aplicarse en su totalidad para proponer una metodología que tenga como eje principal la comprensión empresarial, la continua comunicación con el negocio, el uso de metodologías agiles ,la definición de roles y funciones, el valor agregado de los datos y la aceptación del médico especialista en oncología para la posterior toma de decisiones y futura retroalimentación que ayude a mejorar la agilidad en el diagnóstico del Cáncer de mama para reducir en años posteriores el numero de muertes causadas por esta enfermedad.

Paper Metodología en ciencia de datos para el diagnóstico del Cáncer de mama